

Escola Superior de Saúde Egas Moniz
Mestrado em Biologia Molecular em Saúde



**InDel e SNP: Caracterização Genética das
Populações Imigrantes de Cabo Verde e
Guiné-Bissau a residir na Região de Lisboa**

Rita Antunes dos Reis

Dissertação de Mestrado

Março de 2014

Escola Superior de Saúde Egas Moniz
Mestrado em Biologia Molecular em Saúde



InDel e SNP: Caracterização Genética das Populações Imigrantes de Cabo Verde e Guiné-Bissau a residir na Região de Lisboa

Rita Antunes dos Reis

Candidatura ao grau de Mestre em Biologia Molecular em Saúde

Dissertação orientada pelo Especialista Superior de Medicina Legal António Amorim
e Especialista Superior de Medicina Legal Paulo Dario

Trabalho realizado no Serviço de Genética e Biologia Forenses da Delegação Sul do
Instituto Nacional de Medicina Legal e Ciências Forenses

Março de 2014

Resumo

A República da Guiné-Bissau e a República de Cabo Verde são países africanos cujas histórias se cruzaram aquando da chegada dos navegadores portugueses no século XV. Nos últimos seis séculos, crises políticas, sociais e económicas provocaram ondas de migração dos indivíduos guineenses e cabo-verdianos para Portugal, existindo atualmente e apenas na região de Lisboa cerca de 16 000 imigrantes guineenses e 34 360 imigrantes cabo-verdianos.

Tendo em conta este número elevado de imigrantes e as relações históricas entre Cabo Verde, Guiné-Bissau e Portugal, este estudo teve como objetivo a caracterização genética das populações da Guiné-Bissau e de Cabo Verde a residir na região de Lisboa e a comparação destas com a população de Portugal, com recurso a 30 InDel e 20 SNP. Estes marcadores genéticos apresentam diversas vantagens que os tornam adequados para uso forense, pelo que foram também avaliados para esse fim.

Este estudo incluiu amostras de sangue de 83 indivíduos de origem guineense e de 367 indivíduos de origem cabo-verdiana, que compareceram ao Serviço de Genética e Biologia Forenses da Delegação do Sul do Instituto Nacional de Medicina Legal e Ciências Forenses no âmbito de ações de investigação de parentesco biológico. O estudo dos 30 InDel foi realizado com recurso ao *kit* DIPplex Investigator® (QIAGEN) e dos 20 SNP segundo a metodologia descrita por Dario (2010).

A caracterização genética destas populações permitiu confirmar o contributo do povo guineense na origem do nativo cabo-verdiano, porém não esclarece a influência da população portuguesa na atual constituição genética do povo cabo-verdiano. As diferenças verificadas entre estes povos permitiram a identificação de possíveis marcadores de ancestralidade. Os parâmetros forenses calculados para os 50 marcadores indicam que estes são adequados para uso forense e os resultados obtidos irão poder ser usados na construção de uma base de dados de frequências alélicas, específica para cada uma das populações estudadas.

Palavras-chave: Guiné-Bissau; Cabo Verde; InDel; SNP

Abstract

The Republic of Guinea-Bissau and the Republic of Cape Verde are African countries whose history intersected with the arrival of the Portuguese settlers in the 15th century. Over the past six centuries, political, social and economic crises led to mass migrations of Guinea and Cape Verde individuals to Portugal. Nowadays there are about 16 000 Guinea immigrants and 34 360 Cape Verde immigrants, with fixed residence at Lisbon region.

Given the high number of immigrants and the historical relation between Cape-Verde, Guinea-Bissau and Portugal, this study aimed to characterize the populations of Guinea-Bissau and Cape Verde immigrants in the Lisbon region and to compare these with the population of Portugal, with 30 InDel and 20 SNP markers. These genetic markers have several advantages that make them useful for forensic research, so these were evaluated for that purpose.

This study included blood stains from 83 individuals of Guinea origins and 367 individuals of Cape Verde origins, who attended the Serviço de Genética e Biologia Forenses da Delegação do Sul do Instituto Nacional de Medicina Legal e Ciências Forenses for paternity investigation. Amplification and analysis of the 30 InDel was performed using the Investigator DIPplex[®] kit (QIAGEN) and the 20 SNP according to the methodology described by Dario (2010).

Genetic characterization of populations in study confirms the role of the Guinea people in the origin of native Cape Verde, but does not clarify the influence of the Portuguese people in the current genetic constitution of the Cape Verde population. The differences between these three populations allowed the identification of potential ancestry informative markers. The forensic parameters calculated for the 50 markers indicate that these are suitable for forensic use. The results obtained will be helpful in the construction of an allelic frequencies database, specific for each of the studied population.

Keywords: Guinea-Bissau; Cape Verde; InDel; SNP

Agradecimentos

Gostaria de deixar aqui os meus sinceros agradecimentos a todos os que me acompanharam durante esta etapa da minha vida e que me permitiram concretizar a vontade de trabalhar na área da Genética Forense!

Ao meu orientador Mestre António Amorim, por ter aceite a orientação desta dissertação, por todo o acompanhamento e disponibilidade, por toda a paciência que teve comigo, por todos os conhecimentos transmitidos e pelos inúmeros incentivos que me foi dando no decorrer deste estudo.

Ao meu orientador Mestre Paulo Dario, por ter igualmente aceite a orientação desta dissertação, pela grande paciência que teve para esclarecer todas as dúvidas e dilemas, por todo o acompanhamento e incentivo que me deu e por todos os conhecimentos transmitidos ao longo do trabalho.

À Dr^a. Teresa Ribeiro por me ter recebido no Serviço de Genética e Biologia Forenses do INMLCF-Sul, por tudo o que me ensinou ao longo destes meses (e que foi muito!) e por toda a disponibilidade que teve para comigo.

Ao Sr. Prof. Doutor Jorge Costa Santos por ter aceite a realização desta dissertação de mestrado no INMLCF-Sul.

À coordenação do Mestrado em Biologia Molecular em Saúde, Prof. Dr^a. Alexandra Maia e Silva, por ter aceite este projeto.

Ao restante pessoal do serviço, nomeadamente à Dr^a. Isabel Lucas, à Mestre Heloísa Afonso Costa e à Mestre Cláudia Vieira da Silva por todo o apoio, motivação, boa disposição e principalmente por todos os conselhos que me deram nos momentos mais difíceis!

À Dr^a. Manuela Marques pela grande ajuda que me deu na pesquisa bibliográfica e pela disponibilidade que sempre teve para comigo.

À Rita Oliveira Dario e à Catarina Dourado pela ajuda imprescindível que me deram, quer na realização do trabalho laboratorial, quer na fase de escrita. Sem essa ajuda este trabalho não estava acabado a tempo e horas. Pela amizade e todo o acompanhamento que me deram um especial agradecimento.

Um agradecimento também especial à Filipa Simão, a pessoa que mais me aturou dia e noite nestes últimos meses! Foi um prazer “trabalhar” contigo! Também à Joana Lopes e à Carina Almeida pelas gargalhadas e pelos momentos de descontração que me proporcionaram quando o trabalho era mais que muito!

A todos os meus amigos que passaram fins de semana a ver-me trabalhar e em especial à Sandra Mota, à Rita Baptista, à Inês Clímaco, ao Valter Ramos e ao David Carregueiro, que muitas vezes ouviram os meus desabafos e desesperos, o meu Muito Obrigado pela paciência que tiveram comigo e pela companhia que me fizeram.

Por último e mais importante quero agradecer aos meus pais, Maria Adelaide Reis e Domingos Reis e à minha irmã Susana Reis, pela oportunidade que me deram de realizar este mestrado e pelos sacrifícios que fizeram para que eu pudesse realizar esta dissertação. Não há palavras no Mundo que cheguem para vos agradecer...

A todos, uma vez mais, os meus sinceros agradecimentos, foi um prazer enorme aprender e trabalhar com todos vós!

Índice

ÍNDICE DE TABELAS	IV
ÍNDICE DE FIGURAS E GRÁFICOS	VI
LISTA DE ABREVIATURAS	VIII
1. INTRODUÇÃO	1
1.1. REPÚBLICA DA GUINÉ-BISSAU	2
1.1.1. HISTÓRIA E RELAÇÃO COM PORTUGAL	3
1.1.2. GRUPOS ÉTNICO-LINGÜÍSTICOS	4
1.1.3. DIÁSPORA GUINEENSE	6
1.2. REPÚBLICA DE CABO VERDE	7
1.2.1. DESCOBRIMENTO E COLONIZAÇÃO	8
1.2.2. TRÁFICO DE ESCRAVOS E O PAPEL DE CABO VERDE	10
1.2.3. DIÁSPORA CABO VERDIANA	12
1.3. GENOMA HUMANO	13
1.3.1. ORGANIZAÇÃO DO GENOMA HUMANO	14
1.3.2. MUTAÇÃO E VARIABILIDADE GENÉTICA	15
1.4. GENÉTICA POPULACIONAL	16
1.4.1. EQUILÍBRIO DE HARDY-WEINBERG	17
1.4.2. <i>LINKAGE (DIS)EQUILIBRIUM</i>	20
1.5. MARCADORES GENÉTICOS	21
1.5.1. MARCADORES DE ANCESTRALIDADE	21
1.5.2. AVALIAÇÃO DOS MARCADORES GENÉTICOS	22
1.5.3. MARCADORES GENÉTICOS UTILIZADOS EM GENÉTICA FORENSE	23
1.6. GENÉTICA FORENSE	24
1.7. MARCADORES DO TIPO SNP	25
1.7.1. VANTAGENS E DESVANTAGENS	27
1.7.2. MÉTODOS DE TIPAGEM	28
1.7.3. BASES DE DADOS E PAINÉIS DE SNP	29
1.8. MARCADORES DO TIPO INDEL	33
1.8.1. VANTAGENS E DESVANTAGENS	36
1.8.2. MÉTODOS DE ANÁLISE	37

1.9. OBJETIVOS	40
2. MATERIAIS E MÉTODOS.....	42
2.1. PROCEDIMENTO EXPERIMENTAL	43
2.2. SELEÇÃO DAS AMOSTRAS	44
2.3. EXTRAÇÃO DE ADN	44
2.4. ESTUDO DOS MARCADORES INDEL	45
2.4.1. AMPLIFICAÇÃO POR PCR <i>MULTIPLEX</i>	45
2.4.2. ELETROFORESE CAPILAR DO PRODUTO AMPLIFICADO	47
2.4.3. DETECÇÃO E ANÁLISE DO PRODUTO AMPLIFICADO	47
2.4.4. ALELOS <i>OFF LADDER</i> E <i>DROPOUT</i> DO ALELO D97+	48
2.5. ESTUDO DOS MARCADORES SNP	48
2.5.1. AMPLIFICAÇÃO POR PCR <i>MULTIPLEX</i> E PURIFICAÇÃO DO PRODUTO AMPLIFICADO	48
2.5.2. REACÇÃO DE EXTENSÃO SBE E PURIFICAÇÃO DO PRODUTO SEQUENCIADO	51
2.5.3. ELETROFORESE CAPILAR DO PRODUTO SBE	53
2.5.4. DETECÇÃO E ANÁLISE DO PRODUTO SBE	54
2.6. ANÁLISE ESTATÍSTICA	54
2.7. ANÁLISE DA VARIAÇÃO INTRAPOPULACIONAL E INTERPOPULACIONAL	55
2.8. ANÁLISE FILOGENÉTICA	56
2.9. ANÁLISE DOS PARÂMETROS FORENSES	56
3. RESULTADOS E DISCUSSÃO	58
3.1. ESTUDO DOS MARCADORES INDEL	59
3.1.1. MICROVARIANTES	60
3.1.1.1. <i>Locus</i> HLD97	60
3.1.1.2. <i>Loci</i> HLD99 e HLD84	63
3.2. ESTUDO DOS MARCADORES SNP	64
3.3. CARACTERIZAÇÃO GENÉTICA DAS POPULAÇÕES EM ESTUDO	66
3.3.1. ESTUDO DA POPULAÇÃO GUINEENSE RESIDENTE NA REGIÃO DE LISBOA	67
3.3.2. ESTUDO DA POPULAÇÃO CABO-VERDIANA RESIDENTE NA REGIÃO DE LISBOA	73
3.3.3. COMPARAÇÃO DAS POPULAÇÕES EM ESTUDO	79
3.4. ANÁLISE DA VARIAÇÃO INTRAPOPULACIONAL E INTERPOPULACIONAL	83
3.5. ANÁLISE FILOGENÉTICA	86

3.6. ANÁLISE DE <i>LINKAGE DISEQUILIBRIUM</i>	88
3.7. PARÂMETROS FORENSES	90
3.8. DISCUSSÃO	94
4. CONCLUSÕES.....	100
4.1. ESTUDO DOS MARCADORES INDEL E SNP	101
4.2. ESTUDO DAS POPULAÇÕES IMIGRANTES DA GUINÉ-BISSAU E DE CABO VERDE A RESIDIR NA REGIÃO DE LISBOA	101
5. REFERÊNCIAS BIBLIOGRÁFICAS	103
6. ANEXOS	118
6.1. PERCENTAGENS DE VARIAÇÕES MÁXIMAS E MÍNIMAS DA ANÁLISE INTRA- E INTERPOPULACIONAL	119
6.2. VALORES DE <i>FST</i> OBTIDOS POR <i>LOCUS</i>	120
6.3. DISTÂNCIAS GENÓMICAS ENTRE OS 50 MARCADORES EM ESTUDO	122
6.4. RESULTADOS DA ANÁLISE DE LD ENTRE OS 50 MARCADORES EM ESTUDO	125

Índice de Tabelas

Tabela 1.7.1 – STRs vs SNP.....	28
Tabela 1.7.2 – Marcadores SNP pertencentes ao PCR <i>multiplex</i> Auto 3	32
Tabela 1.7.3 – Marcadores SNP pertencentes ao PCR <i>multiplex</i> Auto 4.	32
Tabela 1.8.1 - Estudos publicados sobre variantes InDel descobertas no genoma humano.....	35
Tabela 1.8.2 – STRs vs InDel.	36
Tabela 1.8.3 – Pannel de 30 marcadores InDel e a Amelogenina incluídos no <i>kit</i> de Investigator®DIPplex (QIAGEN)	38
Tabela 2.4.1 – Constituintes do <i>kit</i> Investigator® DIPplex (<i>QIAGEN</i>).....	45
Tabela 2.4.2 – PCR <i>multiplex</i> . Reagentes e quantidades respectivas, necessários na preparação da mistura de reação para a amplificação dos marcadores InDel.....	46
Tabela 2.4.3 – Condições de temperatura e tempo para que ocorra cada uma das fases necessárias à amplificação dos marcadores InDel e Amelogenina no termociclador.	46
Tabela 2.4.4 – Reagentes necessários para a preparação do produto amplificado param a eletroforese capilar e os respectivos volumes por cada reação.....	47
Tabela 2.5.1 – <i>Primers</i> utilizados para a PCR <i>multiplex</i> dos marcadores correspondentes ao Auto 3.....	49
Tabela 2.5.2 – <i>Primers</i> utilizados para a PCR <i>multiplex</i> dos marcadores correspondentes ao Auto 4.....	49
Tabela 2.5.3 – Reagentes e quantidades respectivas, necessários na preparação da mistura de reação para a amplificação dos marcadores SNP pertencentes ao Auto 3.	50
Tabela 2.5.4 – Reagentes e quantidades respectivas, necessários na preparação da mistura de reação para a amplificação dos marcadores SNP pertencentes ao Auto 4.	50
Tabela 2.5.5 – Condições de temperatura e tempo para que ocorra cada uma das fases necessárias à amplificação dos marcadores SNP no termociclador.	51
Tabela 2.5.6 – <i>Primers</i> utilizados para a reação de extensão, respetiva orientação e SNP detetado com esses mesmos <i>primers</i> para o <i>multiplex</i> correspondente ao Auto 3.	52

Tabela 2.5.7 – <i>Primers</i> utilizados para a reação de extensão, respetiva orientação e SNP detetado com esses mesmos <i>primers</i> para o <i>multiplex</i> correspondente ao Auto 3.	52
Tabela 2.5.8 – Reagentes e respetivos volumes necessários à reação de extensão SBE para o Auto 3, Auto 4 e controlo positivo.....	53
Tabela 2.5.9 – Temperaturas e tempos necessários à reação de extensão SBE.....	53
Tabela 2.5.10 – Reagentes necessários para a preparação do produto SBE para a eletroforese capilar e os respetivos volumes por cada reação.	54
Tabela 3.3.1 – Dados relativos aos 50 <i>loci</i> estudados da população guineense residente na região de Lisboa: número de alelos analisados (2N), frequências alélicas, frequências de heterozigotia (H) e homozigotia (h), heterozigotia observada (Ho) e esperada (He) e o valor de p para o HWE.	68
Tabela 3.3.2 – Dados relativos aos 50 <i>loci</i> estudados da população cabo-verdiana residente na região de Lisboa: número de alelos analisados (2N), frequências alélicas, frequências de heterozigotia (H) e homozigotia (h), heterozigotia observada (Ho) e esperada (He) e o valor de p para o HWE.	74
Tabela 3.5.1 – Matriz de distâncias genéticas F_{st} entre as duas populações em estudo (Guiné-Bissau e Cabo Verde) e outras seis populações caracterizadas com os mesmos marcadores genéticos (30 InDel e 20 SNP).	87
Tabela 3.7.1 – Parâmetros forenses calculados para os 50 marcadores genéticos estudados na população guineense e cabo-verdiana imigrante a residir na região de Lisboa.....	90

Índice de Figuras e Gráficos

Figura 1.1.1 – Mapa da Guiné-Bissau.	2
Figura 1.1.2 – Mapa Etnográfico de Guiné-Bissau. Território da Guiné-Bissau e respetiva ocupação pelos diferentes povos.	5
Figura 1.1.3 – Principais Comunidades imigrantes em Portugal em 2012.	6
Figura 1.2.1 – Mapa da República de Cabo Verde.	7
Figura 1.2.2 – Rotas do tráfico negreiro Ocidental e Árabe entre o século VII e XIX. Para além do tráfico de escravos na região ocidental de África, desde o século VII que este se realiza pelos Árabes.	10
Figura 1.3.1 – Organização do genoma humano dentro da célula: 22 pares de cromossomas autossómicos e um par de cromossomas sexuais dentro do núcleo da célula e centenas de cópias de ADN mitocondrial nas mitocôndrias.	14
Figura 1.7.1 – Exemplo do SNP, identificado por rs1490413, onde ocorre substituição de uma Guanina por uma Adenina, podendo o alelo apresentar duas formas diferentes.	25
Figura 1.7.2 – Reação de Minisequenciação.	29
Figura 1.7.3 – Distribuição dos 20 marcadores SNP ao longo dos diferentes cromossomas. Cada cromossoma está identificado pelo respetivo número em baixo (a azul) e os traços vermelhos representam a localização de cada SNP nos diversos cromossomas.	33
Figura 1.8.1 – Exemplo de um InDel, rs16363, onde se pode observar a variante com a inserção de 5 nucleótidos TGTTT e a variante com a deleção desses mesmos 5 nucleótidos, com alteração do tamanho da sequência.	33
Figura 1.8.2 – Distribuição dos 30 marcadores InDel incluídos no <i>kit</i> de Investigator® DIPplex (QIAGEN) ao longo dos diferentes cromossomas.	39
Figura 2.1.1 – Fluxograma representativo das diversas fases laboratoriais para a realização do estudo dos marcadores InDel e dos marcadores SNP.	43
Figura 3.1.1 – Exemplo de um eletroferograma obtido aquando da deteção e análise dos marcadores do tipo InDel: genótipo de um dos indivíduos em estudo (nº68) do sexo masculino oriundo da Guiné-Bissau.	59
Figura 3.1.2 – Diferenças no tamanho (RFUs) dos picos do <i>locus</i> correspondente à inserção do marcador HLD97 (D97+ marcado a vermelho)	61

Figura 3.1.3 – Microvariantes detetadas aquando da deteção e análise dos marcadores InDel	63
Figura 3.2.1 – Exemplo de um eletroferograma obtido aquando da deteção e análise dos marcadores do tipo SNP: genótipo de um dos indivíduos em estudo (nº278) oriundo de Cabo Verde.	65
Gráfico 3.3.1 – Frequências alélicas para os 20 marcadores SNP em estudo na população guineense imigrante na região de Lisboa.	71
Gráfico 3.3.2 – Heterozigotias observadas e esperadas para os 50 marcadores InDel e SNP estudados na população guineense imigrante residente na região de Lisboa.	72
Gráfico 3.3.3 – Frequências alélicas para os 30 marcadores InDel em estudo na população cabo-verdiana imigrante na região de Lisboa.....	76
Gráfico 3.3.4 – Frequências alélicas para os 20 marcadores SNP em estudo na população cabo-verdiana imigrante na região de Lisboa.....	77
Gráfico 3.3.5 – Heterozigotias observadas e esperadas para os 50 marcadores InDel e SNP estudados na população cabo-verdiana imigrante residente na região de Lisboa.	78
Gráfico 3.3.6 – Frequências alélicas para os <i>loci</i> InDel, obtidas para a população imigrante de origem guineense (GB) e de origem cabo-verdiana (CV) residentes na região de Lisboa: comparação entre os resultados obtidos entre as duas populações.	81
Gráfico 3.3.7 – Frequências alélicas para os <i>loci</i> SNP, obtidas para a população imigrante de origem guineense (GB) e de origem cabo-verdiana (CV) residentes na região de Lisboa: comparação entre os resultados obtidos entre as duas populações.	82
Gráfico 3.3.8 – Heterozigotia Observadas para os 50 <i>loci</i> em estudo na população imigrante de origem guineense e na população imigrante de origem cabo-verdiana.	83
Figura 3.5.1 – Árvore filogenética construída a partir das distâncias genéticas (Fst) demonstradas na Tabela 3.5.1	88

Lista de Abreviaturas

ADN – Ácido Desoxirribonucleico
AIM – *Ancestry informative Marker* - Marcadores Informativos de Ancestralidade
AMOVA – *Analysis of Molecular Variance* – Análise da Variança Molecular
BSA – *Bovine serum Albumin* - Albumina de Soro Bovino
cM - centiMorgans
DIP – *Deletion/Insertion Polymorphism* - Polimorfismo de Deleção/Inserção
dNTPs – Dioxinucleotidos
ddNTP – Dideoxinucleotidos
Fst- Índice de Fixação
HLD – *Human Locus DIP*
He – Heterozigotia esperada
Ho – Heterozigotia observada
HWE – Equilíbrio de Hardy-Weinberg
InDel - Inserção/Delecção
INMLCF – Instituto de Medicina Legal e Ciências Forenses
LE – *Linkage Equilibrium*
LD – *Linkage Disequilibrium*
MgCl₂ – Cloreto de Magnésio
mL – Mililitro
Mpb – Mega pares de bases
ng – Nanogramas
PCR – *Polymerase Chain Reaction* - Reação em Cadeia da Polimerase
PE – Poder de Exclusão
PD – Poder de Discriminação
PIC – *Polymorphism Information Content*
PM – Probabilidade de *Matching*
RFU – *Relative Fluorescence Unit* – Unidade de Fluorescência Relativa
Rpm – Rotações por Minuto
SBE – *Single Base Extension*
SGBF – Serviço de Genética e Biologia Forenses
SNP – *Single Nucleotide Polymorphism*
STR – *Short Tandem Repeats* – Pequenas Repetições em Tandem
SAP – *Shrimp Phosphatase Alkaline* – Fosfatase Alcalina de Camarão
TPI – *Typical Paternity Index* - Índice de Paternidade Típico

1. Introdução

1.1. República da Guiné-Bissau

A República da Guiné-Bissau localiza-se na costa ocidental de África e faz fronteira a norte com o Senegal, a este e a sudeste com a Guiné-Conacri e a sul e a oeste com o Oceano Atlântico. A capital é a cidade de Bissau. Para além do território continental, incorpora o Arquipélago dos Bijagós com oitenta ilhas, separado do continente pelos canais dos rios Geba, Pedro Álvares, Bolama e Canhabaque.

Com uma área de 36 126km², este país é formado no interior por savanas e no litoral por uma planície pantanosa e, por estar entre o Equador e o Trópico de Caranguejo apresenta um clima tropical, caracteristicamente quente e húmido, dividido em duas estações com temperaturas muito elevadas: a das chuvas, de Maio a Novembro e com maior pluviosidade em Julho e Agosto e a estação seca de Dezembro a Abril. Da área total, 22,4% corresponde aos rios e mar e o restante encontra-se dividido em oito regiões e um sector autónomo: Bafatá, Biombo, Bolama, Cacheu, Gabu, Oio, Quinara e Tombali e como sector autónomo Bissau (Figura 1.1.1).



Figura 1.1.1 – Mapa da Guiné-Bissau. Fonte: <http://www.africa-turismo.com/mapas/guine-bissau.htm> [Consultado em 20.01.2014]

Foi uma colónia portuguesa desde o século XV e declarou a sua independência a 24 de Setembro de 1973, tendo esta sido reconhecida por Portugal apenas um ano depois. Atualmente faz parte da Comunidade dos Países de Língua Portuguesa, das Nações Unidas, dos Países Africanos de Língua Oficial Portuguesa e da União Africana.

Relativamente à demografia, a Guiné-Bissau tinha à data dos últimos censos (2009) uma população de 1 520 830 habitantes (Instituto Nacional de Estatística da Guiné-Bissau), a maioria com idades compreendidas entre os 15 e os 65 anos. Composta por diferentes grupos étnicos e com diferentes religiões, este país tem como língua oficial o Português e como não oficial o Crioulo Guineense.

1.1.1. História e Relação com Portugal

A área geográfica que, atualmente, se denomina de Guiné-Bissau, foi descoberta pelo navegador português Nuno Tristão em 1446. A Guiné não compreendia um território bem determinado mas sim uma grande área sem limites, tendo sido chamada durante muito tempo de Rios da Guiné por abranger a área entre o rio Senegal e o norte da Serra Leoa. Só após a divisão do território africano pelos europeus que o colonizavam, é que passou a denominar-se Guiné seguido do nome da soberania, como foi o caso da Guiné-Portuguesa (a atual Guiné-Bissau) e da Guiné-Francesa (a atual Guiné-Conacri ou apenas Guiné). Esta divisão de território e denominação foi estipulada pelo Acto Geral de Berlim de 1885, e foi efetuada de acordo com os interesses da diplomacia europeia, sem respeito pelos limites naturais e pelas divisões políticas e étnicas já existentes e sem considerações pela história comum dos povos (Carreira, 1983a).

Antes da chegada dos europeus e até ao século XVII, quase todo o território da Guiné integrava o reino de Gabu, pertencente ao Império Mali, existente desde 1235. A colonização por parte dos portugueses só teve início um século depois, em 1588. Desde a descoberta até ao início da colonização, a costa da Guiné servia apenas como entreposto comercial, tendo como principal atração para os navios de comércio, os escravos. A colonização ocorreu numa primeira fase apenas nas margens dos rios e no litoral, tendo sido o interior colonizado apenas no século XIX. Contudo, esta colonização decorreu com algumas dificuldades, devidas às investidas dos espanhóis, franceses e holandeses, muito interessados na captura dos escravos.

Também os povos que já habitavam o território dificultaram os avanços e o estabelecimento dos portugueses. A primeira capitania a ser fundada foi a de Cacheu, e só em 1687-1692 é que foi fundada a capitania de Bissau, como fortificação militar e entreposto para o tráfico de escravos, elevada depois a cidade e a capital colonial (PAIGC, 1974).

Os portugueses na Guiné-Bissau foram deixados quase ao abandono pela coroa portuguesa, que monopolizava todo o comércio de mercadorias e de escravos, sobrando apenas a troca de bens para sustento próprio. Isto levou a uma degradação das capitanias e a uma fuga desses portugueses para o interior, que acabavam muitas vezes por aceitar a cultura e os costumes dos diversos povos já existentes, chegando muitas vezes a casar com as mulheres africanas (Carreira, 1983a).

O abuso da captura de escravos, muitas vezes auxiliadas pelos próprios guineenses com maior poder, as guerras territoriais entre os diversos povos, as investidas de outros europeus de monopolizar o comércio e de ocupar o território e a fraca economia, fez com que a Guiné-Bissau e os seus habitantes andassem em constantes guerras. A última guerra civil ocorreu em 1998 e até esse ano foi palco de diversas revoluções e golpes de estado, vivendo em constante crise política até à atualidade (Carreira, 1983a; PAIGC, 1974).

1.1.2. Grupos Étnico-Linguísticos

O território da Guiné-Bissau não é composto por um único povo, mas sim por diversos povos, com religiões e culturas diferentes, que se foram misturando ao longo do tempo. Por essa razão considera-se a existência de diferentes grupos étnico-linguísticos, que ao longo da história de Africa e da Guiné-Bissau foram ocupando territórios e ganhando/perdendo poder. Na Figura 1.1.2 pode-se observar a distribuição dos diferentes povos pelo território da Guiné-Bissau.

A população da Guiné, que teria sido empurrada para o litoral com a chegada dos Mandigas no século XIII e XVI, era composta por diversos grupos étnico-linguísticos: Diola-Balanta, Manjanco e Banhum, Beafadas e Nalús, Bijagós, Cocoli e Padjadincas. O grupo Diola-Balanta vivia organizado em comunidades familiares livres e sem classes hierárquicas, dividindo-se por sua vez em Felupes e Baiotes, que formavam os Diola, e em Balantas, mais numerosos e que se podem dividir em vários ramos aparentados aos Mandigas. O grupo Manjanco e Banhum apresentava

uma organização social hierarquizada e incluía os povos Manjacos, Brames e Papéis que tinham em comum a mesma língua e, os povos Banhum, Cassangas e Coboianas. Estes últimos eram em número reduzido e pouco importantes. O grupo dos Beafadas e Nalús, também pouco importantes para a história do país, teve como influencia os Mandigas, sendo muito semelhantes. Por fim as etnias que se encontravam isoladas e em pequenos grupos, Bijagós que viviam nas ilhas com o mesmo nome, Cocoli e Padjadincas (PAIGC, 1974).

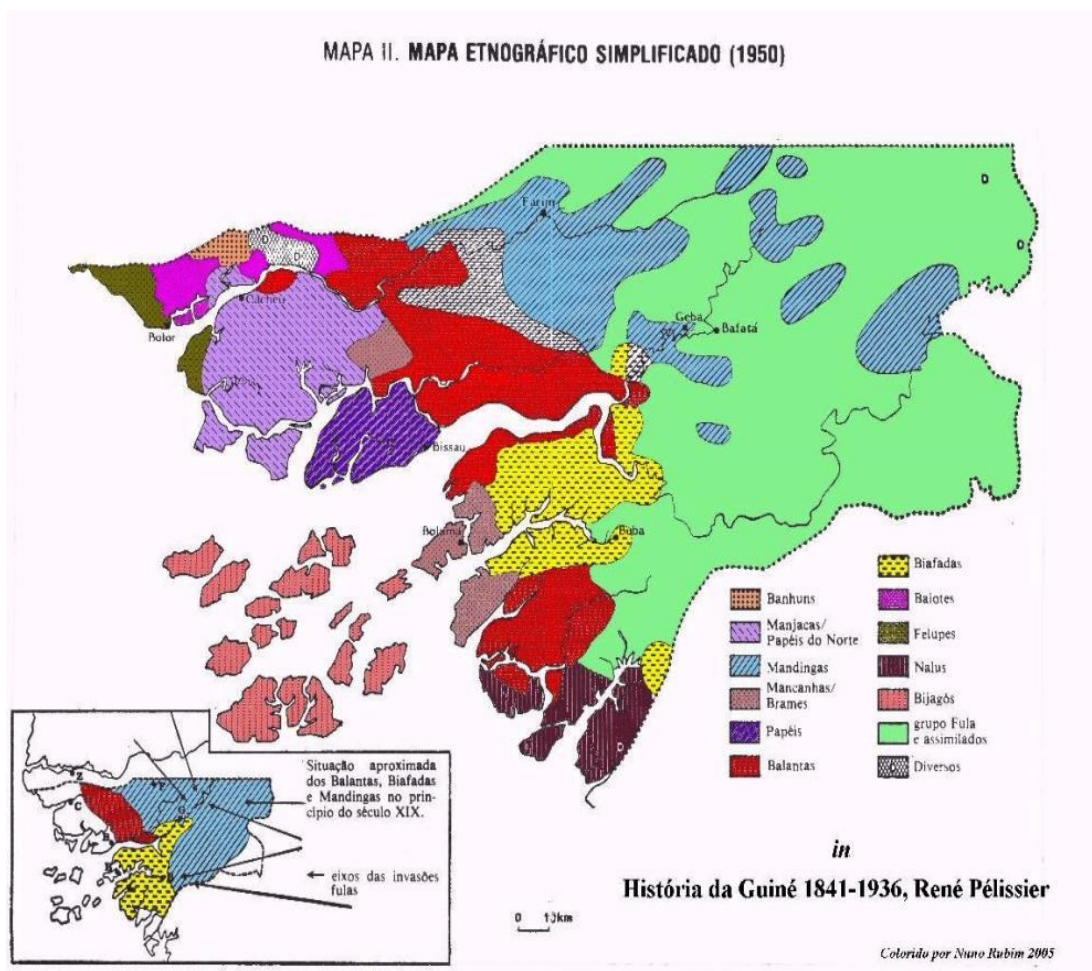


Figura 1.1.2 – Mapa Etnográfico de Guiné-Bissau. Território da Guiné-Bissau e respetiva ocupação pelos diferentes povos. Fonte: blogueforanadaevaotres.blogspot.com [Consultado em 24.01.2014]

O interior da Guiné era composto por Mandigas, vindos do Alto-Niger, que dominavam todo o litoral entre o Gâmbia e o rio Geba (Guiné-Bissau) e que viviam em sistema hierarquizado. Este povo islamita influenciou outros povos em termos de vida social e religião. No interior viviam também os Fulas, que se foram introduzindo na Guiné pouco a pouco, após o século XV (PAIGC, 1974).

1.1.3. Diáspora Guineense

A história da Guiné-Bissau está associada a constantes migrações dos diferentes povos do território. As guerras territoriais, as diversas crises políticas que tanto assombram o país, as crises económicas muitas vezes associadas às políticas, a fome, a falta de trabalho e a pobreza extrema, ditaram ao povo guineense uma constante procura por uma vida melhor e refúgio noutros locais.

A tradição migratória guineense para Portugal tem início por volta dos anos 70, intensificando-se após a independência da Guiné-Bissau e variando nos últimos 40 anos. Entre 1974 e princípio da década de 80, a comunidade guineense em Portugal é composta na sua maioria por luso-guineenses, isto é, indivíduos de nacionalidade portuguesa que nasceram na Guiné-Bissau e por ex-combatentes guineenses, aos quais foram concedidos apoios pelo Estado Português. No fim da década de 80, a crise económica que se instala no país determina um novo fluxo migratório e dez anos depois dá-se um novo fluxo migratório guineense para Portugal, causado pela guerra civil (Abreu, 2011; Cá e Sá, 2012).

Para além de Portugal, o povo guineense tem como principais destinos de emigração o Senegal e a França.

Em Portugal, a comunidade guineense é a sexta maior comunidade de imigrantes, representando 4% do total, como se pode observar na Figura 1.1.3. Esta percentagem refere-se a aproximadamente 17 500 indivíduos guineenses a residir oficialmente em Portugal em 2012 (Serviço de Estrangeiros e Fronteiras, 2013). Desses, 16 119, isto é, cerca de 92%, residem na região Sul de Portugal e praticamente 15 000 residem apenas na região de Lisboa (o que representa 85% da população guineense a residir em Portugal) (Fundação Francisco Manuel dos Santos).

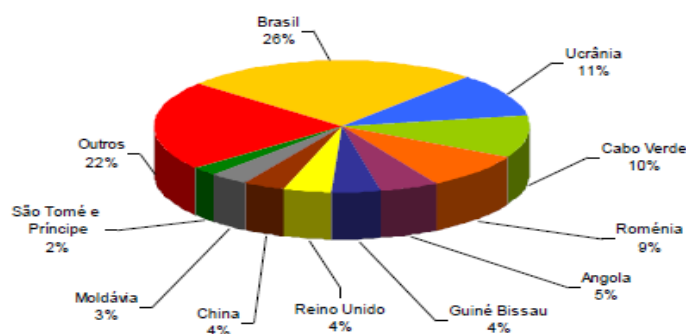


Figura 1.1.3 – Principais Comunidades imigrantes em Portugal em 2012. Fonte: Serviço de Estrangeiros e Fronteiras, 2013

1.2. República de Cabo Verde

A República de Cabo Verde é um país insular africano, de nome originário do cabo costeiro senegalês, Cabo Verde, tendo como língua oficial português e como não oficial crioulo cabo-verdiano. Situado no Oceano Atlântico, a 455Km a oeste da Costa Africana, este arquipélago, de origem vulcânica, apresenta uma extensão total de cerca de 4 033km² e é constituído por dez ilhas e diversos ilhéus, agrupadas segundo a posição relativa aos ventos dominantes de nordeste (Figura 1.2.1). As ilhas de Barlavento incluem as ilhas Santo Antão, São Vicente, Santa Luzia, São Nicolau, Sal e Boa Vista e os ilhéus Branco e Raso. As ilhas de Sotavento incluem as ilhas de Maio, Santiago, Fogo e Brava e os ilhéus do Rombo. É na ilha de Santiago que se localiza a capital do país, Praia, a cidade mais povoada.

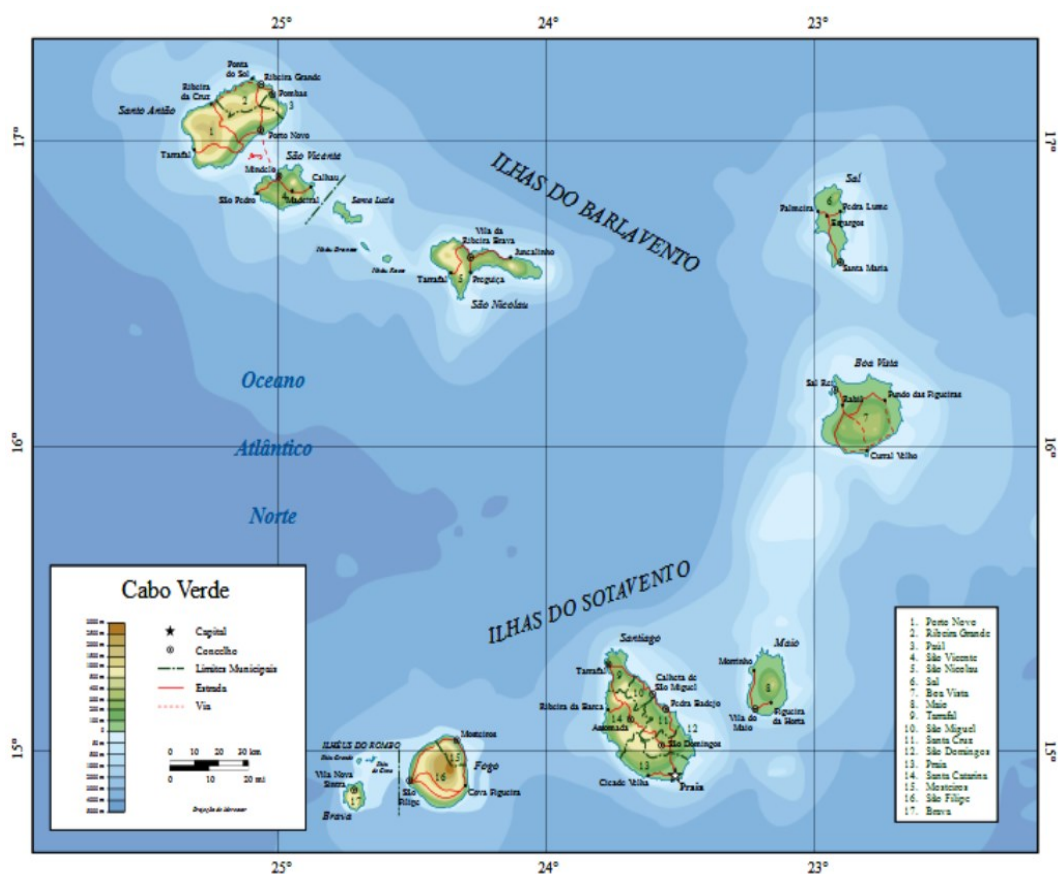


Figura 1.2.1 – Mapa da República de Cabo Verde. Adaptado de <http://commons.wikimedia.org> [Consultado em 19.01.2014]

Localizado na zona subsariana, este país apresenta um clima árido, principalmente nas ilhas com menor relevo, devido à exposição aos ventos quentes e secos do deserto do Sahara. As ilhas com maior relevo apresentam um clima semi-árido e temperado, devido à ação moderadora do oceano e dos ventos alísios, estando

as temperaturas médias anuais entre os 20°C e os 25°C. Cabo Verde apresenta duas estações: das chuvas, também denominada de *as-águas*, de Agosto a Outubro, de elevada irregularidade e de fraca pluviosidade nas ilhas de São Vicente e Sal e, a estação seca, também denominada de *as-secas* ou *tempo das brisas*, de Dezembro a Junho, que é caracterizada por ventos constantes.

Colónia Portuguesa até 5 de Julho de 1975, data em que adquiriu independência, Cabo Verde tem uma população residente (em 2010) de cerca de 491 875 habitantes, o que representa uma densidade populacional de 122 habitantes por quilómetro quadrado, com uma média de idades que ronda os 23-24 anos (Comunidade dos Países de Língua Portuguesa, 2012; INE de Cabo Verde, 2012). Atualmente é integrante da Comunidade dos Países de Língua Portuguesa, das Nações Unidas, dos Países Africanos de Língua Oficial Portuguesa, da União Africana, da Organização Mundial do Comércio e da Comunidade Económica dos Estados da África Ocidental.

1.2.1. Descobrimento e Colonização

As ilhas de Cabo Verde eram já conhecidas por africanos, gregos e geógrafos árabes, muito antes de os portugueses lá chegarem durante a época dos descobrimentos, no século XV. Porém, estas ilhas ficaram esquecidas até que os navegadores Diogo Gomes e Antonio da Noli (genovês), ao serviço do Infante D. Henrique, descobriram as ilhas Sotavento em 1460, referidas no diploma de 3 de Dezembro do mesmo ano. Na Carta de doação de 19 de setembro de 1462, são referidas também as ilhas de Barlavento e dois Ilhéus, achadas nesse mesmo ano, por Diogo Afonso, escudeiro do Infante D. Fernando, este último herdeiro das ilhas do Infante D. Henrique. Através destes dois documentos, consideram-se as datas 1460 e 1462, as oficiais do descobrimento de Cabo Verde (Andrade, 1996; Barcellos, 1899).

Os documentos da época indicam que as ilhas eram desabitadas aquando da chegada dos portugueses, apesar de existirem historiadores que discordam neste assunto e que põem a hipótese de terem existido náufragos a habitar as ilhas (Carreira, 1983b). De qualquer forma, a inexistência de uma população fixa no arquipélago permitiu, a Portugal, iniciar a ocupação das ilhas, enviando em primeiro lugar europeus. Só após avaliadas as possibilidades de fixação e reconhecida a privilegiada situação geográfica, fez-se o povoamento com negros da costa africana.

A ilha de Santiago foi a primeira a ser colonizada, seguida da ilha do Fogo uns anos mais tarde. Só no fim do século XVI e principalmente durante o século XVII é que ocorreu a colonização das restantes ilhas (Barcellos, 1899).

Com o aumento da população na Ilha de Santiago, D. Fernando que tinha ganho o direito de a povoar, dividiu a ilha em duas capitânias (Ribeira Grande, atualmente denominada de Cidade Velha, e Alcatrazes) e cedeu-as aos descobridores do arquipélago como recompensa. As capitânias foram depois repartidas por europeus de estratos sociais com maiores recursos, dando origem ao sistema de morgadios e capelas.

O europeu para além de estar em número reduzido, não tinha resistência física para executar os trabalhos de lavoura e a sua posição de senhor não permitia a execução de tarefas manuais. Para isso recorreu-se a mão-de-obra escrava, a pedido de D. Fernando, através da Carta Régia de 12 de Junho de 1466 que concedia o direito de fazer o comércio de escravos em todas as regiões da costa da Guiné. Em 1472, outra Carta Régia dava aos moradores de Cabo Verde o privilégio de poderem ter escravos, porém sem consentir o comércio destes. O comércio ilegal de escravos começou a fazer-se por livre vontade dos moradores, quando estes se aperceberam da importante fonte de lucro que apresentavam.

Os primeiros grupos a habitar Cabo Verde formavam três classes: brancos, naturais e negros (livres e escravos). Por razões de ordem política, social e religiosa, foram utilizados indivíduos das mais variadas classes sociais na colonização, como homens da nobreza portuguesa, genoveses, exilados, condenados e judeus (Brehm *et al.*, 2002). As mulheres brancas, eram por sua vez muito poucas, muito devido à sua posição social, no caso da nobreza, ou por não serem comuns entre os navegadores, o que levou os homens a criar relações com as escravas negras. Foi a partir destas relações que se originou o nativo cabo-verdiano, também denominado de mulato ou crioulo, com características próprias.

Foi a partir do século XVI que se acentuou a fuga de escravos para o interior das ilhas e até para outras ilhas, contribuindo para o povoamento das mesmas.

1.2.2. Tráfico de Escravos e o Papel de Cabo Verde

Pela sua posição privilegiada entre os três continentes, a ilha de Santiago servia como local de apoio à navegação, principalmente a dirigida ao Golfo da Guiné, costa de Angola, São Tomé, Antilhas e Brasil e de base de apoio, aprovisionamento e de quarentena de navios. Para além disso, esta ilha era um local de concentração de escravos trazidos dos rios da Guiné, para serem ladinizados e exportados. Rapidamente, Cabo Verde tornou-se num ponto central numa complexa rede transatlântica de comércio de escravos (Figura 1.2.2).

O entreposto estabelecido em Santiago resultou da segurança dada pela insularidade, aliada à posição geográfica e à facilidade de escoamento, quer de escravos quer de mercadorias (Carreira, 1983b). A criação de uma base de ladinização no arquipélago, promoveu também o tráfico de escravos. Essa base tinha como objetivo o ensino da catequese, o batismo e a aprendizagem do português, para que o escravo boçal tivesse uma melhor aceitação e cotação nos mercados externos, gerando assim mais lucros.

TRÁFICOS NEGREIROS OCIDENTAIS E ÁRABES, séc. VII - séc. XIX

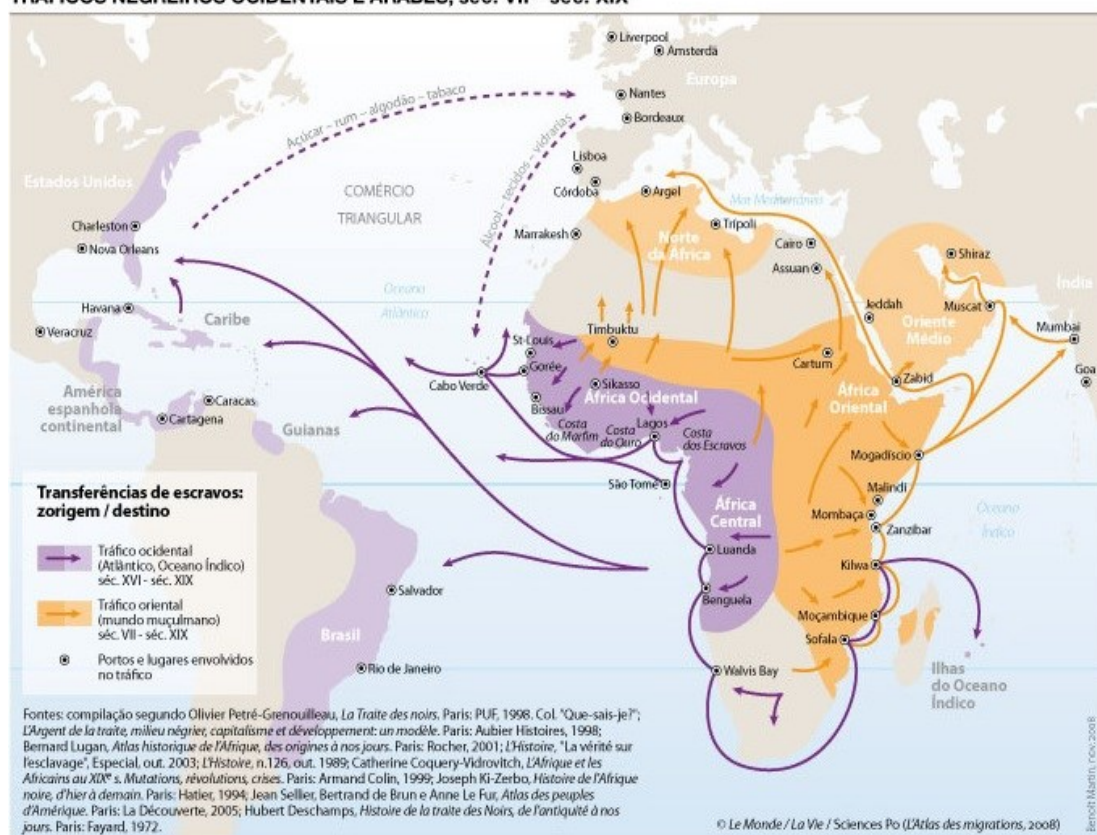


Figura 1.2.2 – Rotas do tráfico negreiro Ocidental e Árabe entre o século VII e XIX. Para além do tráfico de escravos na região ocidental de África, desde o século VII que este se realiza pelos Árabes. Adaptado de <http://www.sciencepro.com.br/> [Consultado em 25.01.2014]

Pelos dados demográficos da época, em 1513, havia na Ribeira Grande (capitania de Santiago) 162 indivíduos: 58 homens brancos, 56 naturais de Portugal, 12 padres, 4 mulheres brancas solteiras, 16 negros e 16 negras. Em 1549, existiam nas duas capitanias um total de 1 200 indivíduos e só em 1582 se incluiu os escravos, perfazendo um total de 15 700 indivíduos, dos quais 13 700 eram escravos, ladinos e boçais. Porém, este número de escravos não refere se era devido à necessidade laboral ou se eram destinados à ladinização e exportação. Pensa-se que apenas um número limitado de escravos seria destinado ao povoamento das ilhas e que parte substancial era vendida para as Antilhas, Brasil e Europa (Carreira, 1983b).

O papel do europeu era orientar e enquadrar os escravos, dirigir os trabalhos e assegurar a disciplina, enquanto o escravo era o empregado na lavoura, no pastoreio, na apanha do algodão e nas tarefas domésticas. O comércio de escravos e todos os géneros associados (mercadorias que serviam como moeda de troca) não era só realizado por portugueses, mas também por espanhóis, ingleses e holandeses, que tinham limitações impostas pela coroa portuguesa. No entanto, qualquer que fosse a origem do navio, os comerciantes tinham de ir em primeiro lugar a Cabo Verde (mais propriamente à Ribeira Grande) fazer os despachos alfandegários e o pagamento dos direitos e só depois poderiam dirigir-se à costa ocidental africana capturar negros escravos, que eram depois taxados, novamente em Cabo Verde. Só então os navios seguiam para o seu destino com a mercadoria. A captura dos negros era muitas vezes realizada pelos autóctones da costa africana, que faziam prisioneiros nas aldeias vizinhas e os trocavam por mercadorias. As zonas que rendiam mais escravos estavam localizadas entre Arguim e o rio Senegal, na Guiné-Bissau, na costa de Mina e na costa de Angola (denominada à época como costa dos Escravos) (Carreira, 1981; Ferronha, 1994).

Com a revogação da lei que obrigava os navios a pagar as taxas em Cabo Verde, (estes passaram a realizar o comércio diretamente da Guiné-Bissau para os seus destinos), Cabo Verde começou a entrar em declínio económico. Isto fez com que os habitantes do arquipélago comessem a negociar, à margem das leis impostas pela coroa portuguesa e fez com que muitos navios não pagassem as devidas taxas, originando todo um comércio de escravos paralelo (Carreira, 1981).

A 24 de Novembro de 1813 foi assinado em Portugal e Inglaterra o primeiro convénio em que ambos se comprometiam a abolir o tráfico de escravos e em 1836,

Marquês Sá da Bandeira publicou o decreto de 10 de Dezembro proibindo a exportação de escravos quer seja por mar ou por terra, em todos os domínios portugueses. Porém o tráfico continuou de forma ilegal. A 28 de Abril de 1858 fixou-se um prazo prorrogável de 20 anos para a abolição total da escravatura e só a 25 de Fevereiro de 1896 foi decretada a extinção da escravatura em todos os domínios portugueses (Ferronha, 1994).

1.2.3. Diáspora Cabo Verdiana

A história das ilhas de Cabo Verde está marcada por fenómenos que explicam os movimentos migratórios da população cabo-verdiana. O indivíduo cabo-verdiano, pelas secas, fomes, defeituosa repartição das terras e desequilíbrio população/produção é um eterno emigrante em busca do que a terra lhe nega sistematicamente (Carreira, 1983b).

À época da colonização, a economia das ilhas de Santiago e do Fogo resumia-se ao cultivo de milhos e tubérculos para o sustento dos escravos e abastecimentos dos navios, na criação de gado, leite e derivados destes e na apanha de sal, urzela, urucú e anil vegetal. As indústrias eram poucas. O comércio dos produtos e escravos destinava-se quase exclusivamente à coroa portuguesa, pelo que pouco ou nada sobrava para o povo. Para além disso, o pouco comércio que o povo podia realizar era baseado na troca de bens, o que limitava a circulação de dinheiro e originava uma economia pobre, ainda hoje característica de Cabo Verde.

Também o clima foi, de certa forma, responsável por vagas de emigração do povo cabo-verdiano. A pluviosidade irregular e os anos de seca criaram problemas nas plantações que serviam de sustento ao povo, aliada ao crescimento da população (que oscilava consoante as conveniências de reposição de escravos) e consequente ocupação dos poucos terrenos férteis, e aos ventos de leste que ciclicamente fustigam o arquipélago, levaram a diversas crises de alimentos e água, que levaram por sua vez a números elevados de mortalidades, miséria e emigração.

A longa história migratória deste povo começa no fim do século XVII e princípio do século XVIII, quando começaram a partir para a América. Depois desse período existiram outros três, caracterizados por grandes fluxos migratórios. O primeiro entre 1900 e 1926, para os Estados Unidos. O segundo entre 1927 e 1945, com destino aos países da América Latina (Brasil e Argentina) e África (Senegal,

Guiné-Bissau, São Tomé e Príncipe e Angola). Foi nesta altura que Portugal começou a destacar-se como destino dos emigrantes cabo-verdianos. O terceiro ocorreu entre 1946 e 1973, onde começaram a ir para a Europa. Estes países continuam a ser os destinos de eleição para estes emigrantes (Carvalho, 2010).

Atualmente, de acordo o governo de Cabo Verde, a população emigrada cabo-verdiana é superior à residente no país. Só em Portugal existem (dados de 2012), cerca de 42 388 imigrantes oriundos do arquipélago, sendo que desses, 37 667 são residentes na região Sul de Portugal (compreende Lisboa, Alentejo e Algarve) e 34 364 apenas na região de Lisboa, o que representa respetivamente 89% e 81% da população imigrante cabo-verdiana em Portugal (Fundação Francisco Manuel dos Santos). Os imigrantes oriundos de Cabo Verde a residir em Portugal representam 10,3% (em 2012) de toda a população imigrante em Portugal, sendo a terceira maior comunidade (Figura 1.1.3) (Serviço de Estrangeiros e Fronteiras, 2013).

1.3. Genoma Humano

O genoma humano pode ser definido como toda a informação genética existente numa célula humana, estando essa informação genética contida numa molécula denominada de ácido desoxirribonucleico (ADN).

O ADN é uma macromolécula composta por quatro unidades, compostas por sua vez por três subunidades: um açúcar desoxirribose, uma base azotada e um grupo fosfato. As quatro bases (unidades) constituintes do ADN dividem-se em dois tipos, em purinas (adenina e guanina) e em pirimidinas (timina e citosina). Estas bases juntamente com o açúcar e o grupo fosfato formam os nucleótidos (Nussbaum *et al.*, 2007). A informação contida no ADN descreve-se em quatro letras, representativas de cada nucleótido constituinte, isto é, por, A (adenina), C (citosina), G (guanina) e T (timina). É com estas quatro letras e com as combinações possíveis destas que assenta toda a variabilidade genética (Butler, 2005a).

Dentro de uma célula, a informação genética está armazenada dentro do núcleo (ADN nuclear) e das mitocôndrias (ADN mitocondrial) (Figura 1.3.1).

O ADN mitocondrial (ADNmt), que representa uma pequena percentagem da totalidade do genoma, é uma pequena molécula circular, de cadeia dupla, que contém cerca de 16 569 pb, dos quais 1 122pb pertencem à região controlo não codificante,

onde existem três regiões hipervariáveis. Na cadeia dupla, uma das cadeias é denominada leve e a outra pesada, porque apresenta, esta última, maior conteúdo em C-G. Outra característica desta molécula é o facto de a sua transmissão ocorrer apenas por via materna (Butler, 2005a).

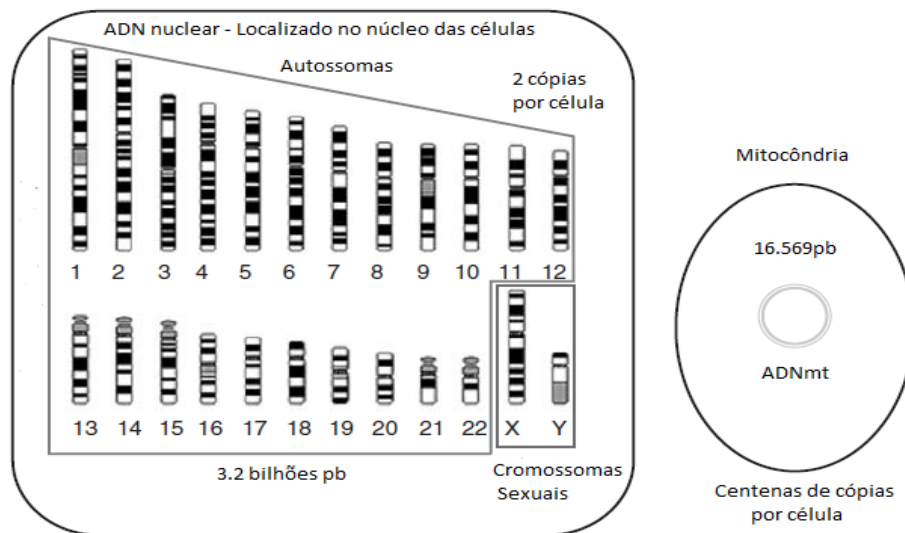


Figura 1.3.1 – Organização do genoma humano dentro da célula: 22 pares de cromossomas autossómicos e um par de cromossomas sexuais dentro do núcleo da célula e centenas de cópias de ADN mitocondrial nas mitocôndrias. Adaptado de Butler (2010a)

A maior parte do genoma humano é constituído pelo ADN nuclear, que se encontra em 46 cromossomas, 22 pares autossómicos e 2 cromossomas sexuais (X e Y), que variam em tamanho e que apresentam cerca de 3 bilhões de pares de bases no total (Makałowski, 2001). A transmissão do ADN nuclear ocorre por via materna e por via paterna, onde cada progenitor transmite metade da informação genética.

1.3.1. Organização do genoma Humano

O material genético contido nos cromossomas é composto por regiões codificantes e por regiões não codificantes. As regiões codificantes referem-se aos genes que contêm toda a informação necessária para a produção de proteínas e que são constituídos por intrões e por exões. Os genes representam apenas cerca de 5% do genoma humano, sendo que os restantes 95% referem-se às regiões não codificantes, que podem estar ou não relacionadas com a produção de proteínas. É nestas regiões não codificantes que se encontram os marcadores genéticos utilizados

em genética forense, quer estejam entre genes (ADN intergénico) ou inseridos nos genes (intragénicos) (Butler, 2005a).

As sequências de ADN podem ser classificadas em sequências de cópia única, que como o nome indica apresentam apenas uma cópia ao longo de todo o genoma, e em sequências repetitivas, que podem ser distinguidas pela sua localização, isto é, se se encontram dispersas pelo genoma ou agrupadas em locais específicos. Estas últimas consistem em grupos de pequenas repetições organizadas em tandem, que variam no número de unidades de repetição e no tamanho (Nussbaum *et al.*, 2007b).

As sequências repetitivas em tandem, também denominadas de ADN satélite, e dependendo do seu tamanho, podem ser: macrossatélites, que podem ter até centenas de quilobases; minissatélites com um tamanho menor que 1-15kb; ou microsatélites que apresentam cerca de 4pb ou menos (Kass e Batzer, 2001). É nestas sequências repetitivas em tandem que se encontram os marcadores genéticos mais comumente utilizados em genética forense, os STR (do inglês *Short Tandem Repeats*) e mais recentemente os miniSTR (Goodwin *et al.*, 2007a).

1.3.2. Mutação e Variabilidade Genética

O ADN é uma molécula que tende a sofrer alterações, principalmente durante o processo de divisão celular. Contudo, existem diversos mecanismos celulares que reparam esses erros introduzidos no ADN. Nem sempre os mecanismos de reparação são eficazes e as alterações permanecem na sequência de ADN, passando a denominar-se de mutações. Uma mutação é então definida como qualquer alteração permanente na sequência nucleotídica ou no rearranjo do ADN (Nussbaum *et al.*, 2007a).

As mutações podem alterar o número de cromossomas, a estrutura dos cromossomas ou a sequência genómica. Podem também ocorrer nas regiões codificantes do genoma, levando a alterações nos genes e/ou nas suas funções, e nas regiões não codificantes, muitas vezes sem manifestações (Martins e Carvalho, 2008; Nussbaum *et al.*, 2007a).

As mutações que alteram a sequência genómica podem ser substituições de nucleótidos, ou inserções, deleções, duplicações e translocações de nucleótidos. Relativamente às substituições, estas podem ser de dois tipos: transições, onde ocorre

substituição de uma purina por outra purina ou de uma pirimidina por outra pirimidina e; transversões, onde ocorre substituição de uma purina por uma pirimidina ou vice-versa. As deleções e inserções podem ocorrer em mais do que um nucleótido e podem, levar a alterações a nível funcional por alteração da grelha de leitura (mutações *frameshift*). Estas alterações dão origem a variantes alélicas, que quando presentes em mais de 1% da população são denominadas de polimorfismo. Quando as mutações ocorrem em células somáticas, vão dar origem a um conjunto de células mutantes idênticas, que afetam apenas o indivíduo onde surgiu, sem transmissão à descendência. Quando ocorrem em células da linha germinativa, podem ser transmitidas à descendência, sendo estas as de interesse para a análise forense (Martins e Carvalho, 2008; Nussbaum *et al.*, 2007a).

É com base na análise das mutações que foram ocorrendo ao longo do tempo (na região não codificante do genoma) que se constata a grande variabilidade genética entre indivíduos e entre populações. Essas mutações por não sofrerem pressão seletiva (como são neutras, não apresentam manifestações funcionais) são fixadas no genoma e passam à descendência (Martins e Carvalho, 2008).

A variabilidade genética existente entre indivíduos está presente apenas em 1% do genoma humano, sendo a sequência genómica e a informação contida neste, idêntica nos restantes 99%. As regiões variáveis encontram-se bem caracterizadas e são estas que são utilizadas em estudos de genética forense (Goodwin *et al.*, 2007a). Porém, é necessário que a variabilidade existente num *locus* seja estável o suficiente para ser passada à descendência mas não tão estável que não dê origem a novos alelos e não permita que o *locus* seja muito informativo.

Para além das mutações, existe outro fator que confere variabilidade entre indivíduos: a recombinação. Este é um processo onde ocorre uma troca aleatória entre segmentos de cromossomas homólogos durante a meiose e que dará origem a gâmetas diferentes. A combinação destes gâmetas de cada um dos progenitores é o que permite que cada indivíduo seja único (Butler, 2010b).

1.4. Genética Populacional

O conhecimento da constituição e da variabilidade genética de grupos específicos de indivíduos, ou seja, de populações, é uma ferramenta muito útil em Genética Forense, pois o estudo das frequências e das combinações alélicas de

populações permite a construção de bases de dados. Estas por sua vez permitem inferir acerca das frequências dos perfis de ADN obtidos em qualquer laboratório forense (Butler, 2005b).

Uma população, em Genética Forense, refere-se a um grupo de indivíduos com a mesma ancestralidade (mesma origem geográfica) (Goodwin *et al.*, 2007b), ou com a mesma origem étnica, existindo três grandes grupos populacionais humanos: Europeia, Africano e Asiático.

O estudo genético das populações baseia-se no estudo das frequências alélicas e genotípicas em diferentes *loci* numa população e na forma como estas se mantêm, ou se alteram, consoante o tempo, visando a quantificação das variações observadas dentro de uma população específica ou entre populações diferentes (Butler, 2005b; Nussbaum *et al.*, 2007a). Os estudos genéticos populacionais não têm apenas em conta os fatores genéticos de variação, como as mutações e os padrões de cruzamento, têm também em conta fatores sociais e ambientais como a migração e a seleção natural.

A construção de uma base de dados de frequências alélicas é realizada pela contagem das ocorrências dos alelos numa determinada população, logo, quanto maior for o número de indivíduos integrantes numa base de dados, mais representativa e precisa esta será, contudo é necessário que os indivíduos que a compõem não estejam relacionados diretamente. O uso das bases de dados, juntamente com a análise do Equilíbrio de Hardy-Weinberg (HWE), permite ao especialista forense o cálculo da frequência de um perfil genético pela multiplicação das frequências genotípicas em cada *locus* e, só é possível pelo facto de a hereditariedade dos alelos em cada *locus* ser independente de outros *loci* (*linkage equilibrium*) (Goodwin *et al.*, 2007c).

1.4.1. Equilíbrio de Hardy-Weinberg

O princípio de Hardy-Weinberg permite uma representação matemática simples das relações das frequências alélicas e genotípicas numa população em que estas se encontrem em equilíbrio. Este princípio diz que numa população infinitamente grande, que se cruza ao acaso, as frequências genotípicas num *locus* mantêm-se constantes. Quando uma população está em equilíbrio de Hardy-

Weinberg (HWE), as frequências genóticas das gerações seguintes, podem ser previstas pelas frequências alélicas.

As frequências genóticas referem-se ao número de indivíduos com determinado genótipo dividido pelo número total de indivíduos estudados. As frequências alélicas referem-se ao número de cópias de um alelo na população estudada dividido pelo número total de alelos observados nessa mesma população. Estas são herdadas de forma Mendeliana e seguem um padrão previsível, isto é, se dois alelos, A e a , apresentam frequências p e q , respetivamente, então um genótipo homozigótico AA deverá ocorrer a p^2 (ou aa a q^2) e um genótipo heterozigótico deverá ocorrer a $2pq$. Quando uma população apresenta frequências genóticas observadas semelhantes às frequências genóticas esperadas (calculadas através das frequências alélicas observadas), diz-se que está em HWE. Outra implicação de uma população estar em HWE é que quando um alelo é raro, este vai existir mais na forma heterozigótica do que homozigótica.

Para uma população ser ideal e estar em HWE, deve cumprir cinco pressupostos: deve ser infinitamente grande, os cruzamentos devem ocorrer ao acaso e não deve estar sujeita a efeitos de migração, seleção natural e a mutações. Como é de esperar, nenhuma população humana respeita estas condições e por isso podem não estar em HWE. Os efeitos de migração, seleção e de mutação são também fatores importantes quando se pretende entender a evolução da espécie humana e das populações que a constituem.

Uma das consequências de ter uma população com um tamanho finito (como por exemplo as populações geograficamente isoladas) é a ocorrência de um fenómeno denominado de deriva genética. Este caracteriza-se pela alteração das frequências alélicas (que podem aumentar ou diminuir) entre as diferentes gerações. Quando se fala de um estudo populacional, o número de indivíduos estudados para inferir acerca das frequências de uma população deve ser o maior possível, para evitar que este efeito se pronuncie de forma significativa. Para além desta consequência, o tamanho da população pode afetar também a forma como ocorrem os cruzamentos entre indivíduos, fazendo com que estes não ocorram ao acaso, contrariamente ao que é suposto para uma população estar em HWE e, fazendo também com que os indivíduos muitas vezes partilhem um ancestral em comum, o

que leva por sua vez um decréscimo das frequências heterozigóticas e um aumento das homozigóticas.

Em qualquer população humana existe uma seleção do parceiro pelas características fenotípicas mais apelativas, o que irá influenciar a forma como ocorrem os cruzamentos entre indivíduos e que será mais acentuado quanto menor for uma população, pois existem muitas vezes relações de consanguinidade entre indivíduos que vão aumentar as frequências de homozigóticos nessa população (Pasternak, 2005). Porém quando se fala em Genética Forense, os *loci* estudados não conferem qualquer característica fenotípica logo, a seleção do parceiro para criar descendência não é influenciada pelos genótipos desses *loci*, logo, considera-se que os cruzamentos entre indivíduos ocorrem sempre ao acaso, sem influenciar o HWE.

Ao longo da história, a população humana esteve sempre sujeita ao efeito das migrações. Este efeito pode provocar desvios no HWE pois levam a alterações no conjunto genético de uma população onde ocorreu imigração, por um fenómeno denominado de fluxo génico. Este fenómeno tem como consequência a mistura genética de duas ou mais populações com composição genética diferente, sendo o impacto provocado dependente do número de indivíduos que entram na população, do tamanho da população nativa e das diferenças das frequências alélicas entre a população de origem dos indivíduos e a população que os recebe. Os indivíduos imigrantes podem também inserir novos alelos numa população, provocar uma redução das frequências homozigóticas e igualar as diferenças genéticas entre as diferentes populações (Pasternak, 2005). O HWE, desviado pelo fluxo génico, pode ser restaurado ao fim de uma geração, ficando a nova população em equilíbrio.

Na seleção natural, por ocorrer uma pressão seletiva que mantém as alterações genéticas vantajosas e elimina as desvantajosas, pode também ser alterado o HWE numa população, pois há uma tendência para ocorrer um aumento das frequências de alelos vantajosos e ao mesmo tempo um decréscimo das frequências de alelos menos vantajosos para essa população. Contudo, ao estudar as frequências alélicas, em Genética Forense, não são considerados *loci* que confirmam características que possam ser mais ou menos vantajosas e que possam vir a alterar as frequências alélicas das gerações seguintes, portanto nestes casos, considera-se que este pressuposto não altera o HWE numa população.

Quanto à última condição, se uma população estiver sujeita a efeitos de mutação nos *loci* estudados, esta irá sofrer alterações nas frequências alélicas que podem levar a um desvio do HWE. Este desvio é mais notório quando as taxas de mutação são elevadas e quando ocorre a introdução de novos alelos no conjunto genético de uma população. Se esses novos alelos se fixarem na população, após algumas gerações (tal como ocorre aquando das migrações) o HWE é restabelecido.

1.4.2. *Linkage (dis)equilibrium*

No genoma humano, existem segmentos de ADN num mesmo cromossoma que se encontram em associação e que são transmitidos à descendência em bloco. Esses segmentos podem conter diversos *loci* que irão então ser transmitidos em associação, isto é, em *linkage*. O estado que descreve as relações entre alelos em diferentes *loci* pode ser denominado de *linkage equilibrium* (LE), caso os alelos sejam segregados de forma independente e aleatória ou de *linkage disequilibrium* (LD), caso os alelos sejam transmitidos em associação não aleatória (Buckleton, 2005).

Tendo em conta que o cálculo das frequências alélicas e genotípicas têm como base as duas leis de Mendel, a lei da segregação e a lei da segregação independente, isto é, com base no facto de que os alelos de cada par se separam durante a meiose, recebendo cada gâmeta apenas um dos alelos e que a transmissão de um alelo não influencia a transmissão de outro, se ocorrer LD entre alelos, então estes não podem ser, estatisticamente, considerados como independentes (Buckleton, 2005; Fondevila *et al.*, 2011c).

Para a obtenção de um perfil genético são estudados um elevado número de marcadores, o que aumenta a probabilidade de diferentes *loci* se encontrarem numa posição genómica relativamente próxima, que é uma das causas da ocorrência de LD. Também a taxa de recombinação pode influenciar a ocorrência de LD (Fondevila *et al.*, 2011b). Se esta taxa for pequena, significa que ocorre pouca recombinação entre cromossomas homólogos durante a formação dos gâmetas, logo existem alelos que são transmitidos à descendência em bloco e novamente não podem ser tratados como independentes. Para assegurar uma independência entre alelos, estes devem estar localizados preferencialmente em diferentes braços cromossómicos e separados por

pelo menos 50 cM, o que significa que existe uma probabilidade de 50% de ocorrer recombinação entre dois marcadores (Westen *et al.*, 2012).

O efeito de *linkage* pode ser alterado por diversos fatores moleculares, entre os quais as mutações que podem reduzir o nível de associação entre diferentes *loci*, e também pelos fatores populacionais que afetam o HWE. Por esta razão, a extensão de LD pode apresentar variações entre diferentes populações e entre diferentes regiões do genoma (Hinds *et al.*, 2005; Kauppi, 2003).

1.5. Marcadores Genéticos

Os marcadores genéticos são sequências de ADN conhecidas, transmissíveis à descendência e de localização conhecida, que permitem individualizar um indivíduo.

Como já foi referido, os marcadores polimórficos ou marcadores genéticos utilizados em Genética Forense, encontram-se nas regiões não codificantes no genoma. A localização cromossômica de um marcador é denominada de *locus*, sendo o termo alelo utilizado para possíveis alternativas do *locus* genético. A caracterização de todos os alelos presentes no *locus* é denominada de genótipo e a combinação da informação de diferentes *loci* é denominada de perfil de ADN ou perfil genético (Butler, 2010b).

Para serem considerados bons marcadores, estes têm de cumprir algumas características, entre as quais: níveis de heterozigotia elevados; localização em cromossomas distintos; baixas taxas de mutação; com amplicões de pequeno tamanho; não sofrer qualquer pressão seletiva; facilidade e reprodutibilidade técnica (Goodwin *et al.*, 2007a; Souto, 2010).

1.5.1. Marcadores de Ancestralidade

O significado de população é muitas vezes ambíguo, sendo irregularmente definido pela etnia, geografia, cultura, religião, entre outros. Rosenberg e colaboradores, em 2002, verificaram que dentro de uma população, as diferenças entre indivíduos representavam 93-95% e entre regiões apenas 3.6% da variação genética total.

A análise da estrutura genética de uma população pode assim, por diferenças nas frequências alélicas, identificar a origem geográfica (principalmente continental) da mesma (Kosoy *et al.*, 2009; Rosenberg *et al.*, 2005). Mais especificamente, pode identificar a origem de um indivíduo. Para além disso, permite estudos de evolução e inferir acerca das subestruturas da população em estudo (Yang *et al.*, 2005).

Vários painéis de marcadores de ancestralidade (AIM do inglês *Ancestry Informative Marker*) foram desenvolvidos para populações hispânicas, afro-americanas, europeias e para determinação a um nível global (Nievergelt *et al.*, 2013).

Os principais marcadores AIM utilizados têm como base o cromossoma Y e o ADNmt, que são marcadores de linhagem paterna e materna, respetivamente. Porém, têm vindo a ser estudados outros marcadores genéticos, que possam ser utilizados para fornecer indicações acerca da origem geográfica de indivíduos, útil quando numa análise forense, nada se sabe acerca do dador da amostra (Lao *et al.*, 2006; Phillips *et al.*, 2007).

1.5.2. Avaliação dos marcadores genéticos

Para utilizar um conjunto de marcadores genéticos em investigações forenses, é necessário uma avaliação estatística prévia de cada marcador e de todo o conjunto desses marcadores. Essa avaliação inclui o cálculo de diversos parâmetros, como o nível de heterozigotia (H), a probabilidade de *matching* (PM), o poder de discriminação (PD), o poder de exclusão (PE), o PIC (do inglês *polymorphism information content*) e o índice de paternidade típico (TPI do inglês *typical paternity index*).

O nível de heterozigotia de um marcador genético indica a proporção de indivíduos heterozigóticos para esse marcador numa população. Quanto maior for esse nível, maior é a diversidade de alelos existentes numa população e logo menor é a probabilidade de haver correspondência entre duas amostras ao acaso.

A probabilidade de *matching* ou de identidade indica a probabilidade de dois indivíduos selecionados ao acaso apresentarem um genótipo idêntico num determinado *locus*. O poder de discriminação, é o contrário da PM, e indica a probabilidade de dois indivíduos selecionados ao acaso apresentarem genótipos diferentes.

Quanto ao poder de exclusão, este parâmetro refere-se à probabilidade de um qualquer suposto pai ser excluído, considerando todas as possibilidades entre a mãe e o filho.

O *polymorphism information content* reflete a probabilidade de um dado descendente de um progenitor com um alelo raro num *locus* vir a permitir a dedução do genótipo parental nesse *locus*.

Por último, o índice de paternidade típico reflete a probabilidade de que o indivíduo testado seja o verdadeiro pai biológico em vez de outro indivíduo não testado, para um dado *locus*, assim, quanto maior o TPI, maior é a probabilidade de o indivíduo ser o pai biológico (Butler, 2005b).

1.5.3. Marcadores Genéticos utilizados em Genética Forense

Existem diversos marcadores genéticos utilizados em genética forense, entre eles, STR, miniSTR, InDel (inserção/deleção) e SNP (do inglês *Single Nucleotide Polymorphism*).

Os STR são muito abundantes no genoma humano e apresentam um tamanho de unidades de repetição variável, sendo os mais utilizados os tetranucleotídicos pelo facto de possuírem alelos com tamanhos muito próximos. Isto permite uma análise numa única reação, produtos de PCR de tamanho pequeno, redução da amplificação preferencial de alelos com pequeno tamanho e a diminuição de fenómenos que dificultam a interpretação de resultados. Existem vários tipos de marcadores STR que podem ser analisados, como os STR autossómicos e os STR dos cromossomas sexuais (Pinheiro, 2008). Estes marcadores apresentam sequências de ADN entre 2pb e 7pb que se repetem entre 2 a 50 vezes e cujos alelos apresentam tamanhos entre 50 a 300-500pb (Butler, 2010b; Goodwin *et al.*, 2007a; Pinheiro, 2008).

Relativamente aos miniSTR, a principal característica vantajosa deste tipo de marcadores é o facto de originarem produtos de amplificação de tamanho inferior a 150pb, o que os torna muito úteis na análise de pequenas quantidades de ADN e de ADN degradado. Tal como os STR, estes podem ser autossómicos ou de cromossomas sexuais (Butler, 2010b; Pinheiro, 2008).

Na análise destes marcadores, as variações alélicas observadas são devidas aos diferentes números de unidades de repetição, que resultam em alelos de

diferentes tamanhos. Estes polimorfismos são denominados de polimorfismos de comprimento (Goodwin *et al.*, 2007a).

Existem outras duas classes de marcadores que têm vindo, nos últimos anos, a ganhar interesse por parte dos geneticistas forenses, sendo esses os SNP e os InDel, referidos em maior pormenor nos capítulos seguintes.

1.6. Genética Forense

Em 1985, Jeffreys e colaboradores, relataram a existência de regiões no genoma humano que se repetiam várias vezes, e que o número de repetições dessas regiões variava de indivíduo para indivíduo, sugerindo que estas podiam ser utilizadas em casos forenses (Butler, 2005c; Jeffreys *et al.*, 1985a; Pinheiro, 2010). Foi nesse mesmo ano que este autor descreveu o *DNA fingerprint*, que permitia detetar diversos *loci* minissatélites por restrição enzimática e hibridação com sondas específicas para sequências das regiões repetitivas (Gill *et al.*, 1985).

O primeiro caso onde foi utilizada a nova técnica de tipagem de ADN com marcadores microsatélites foi um caso de imigração, em 1985, resolvido por Alec Jeffreys. Tratava-se de um rapaz ganês nascido no Reino Unido, que aquando da separação dos pais emigrou para o Gana para se juntar ao pai. Quando voltou ao Reino Unido, para se reunir com a mãe, o irmão e as duas irmãs, viu a sua residência recusada. Havia suspeitas de que não era o mesmo indivíduo e que teria havido uma substituição por um primo ou por outro qualquer rapaz ganês. Análises aos marcadores genéticos convencionais, isto é, testes serológicos, demonstraram que a mãe e o indivíduo em questão eram relacionados, mas não esclarecia se era o filho ou um sobrinho. Foi então que foi feito um pedido para a realização de uma análise ao ADN, após notícia sobre o assunto, para determinar a maternidade do rapaz em questão. Apesar de a técnica ainda não estar aperfeiçoada, Jeffreys aceitou analisar as amostras da família que estava disponível, ou seja, a mãe do rapaz, o irmão e as duas irmãs, utilizando o método desenvolvido pelo próprio. Por comparação com os supostos irmãos e com a mãe e, com uma análise cuidadosa de todas as hipóteses, Jeffreys verificou que a mãe era a verdadeira mãe do indivíduo em questão, tendo-lhe sido concedida residência no Reino Unido (Aronson, 2005; Jeffreys *et al.*, 1985a).

Antes destas descobertas, a genética forense tinha como base a análise de marcadores serológicos, como os grupos sanguíneos, os antígenos de leucócitos humanos, as proteínas e as enzimas, sendo o estudo destes marcadores realizado com base em métodos imunológicos e/ou eletroforéticos. Foi com a descoberta da PCR (do inglês *Polymerase Chain Reaction*), na década de 80, por Kary Mullis, que esta área se revolucionou. Também a introdução de novas tecnologias como o desenvolvimento dos termocicladores e os sequenciadores automáticos e o desenvolvimento de métodos de extração e de *kits* comerciais de tipagem de ADN, levaram a um grande avanço na resolução de casos forenses (Pinheiro, 2008, 2010). Três décadas após as grandes descobertas, esta área ainda se encontra em constante evolução e inovação, quer nas tecnologias utilizadas, quer nos marcadores estudados.

Atualmente, a genética forense é uma área da genética, que com recurso a amostras biológicas e técnicas de biologia molecular, se dedica à resolução de casos de investigação de parentesco biológico, de migração, de identificação individual e de criminalística biológica, sempre por comparação dos perfis genéticos obtidos. A identificação genética é um processo que permite estabelecer a identidade de um indivíduo através de um conjunto de caracteres que o individualizam, sendo esses caracteres um pequeno conjunto de variações genéticas no genoma humano (Butler, 2010b; Pinheiro, 2008).

1.7. Marcadores do tipo SNP

No genoma humano existem variações que ocorrem por substituição de um só nucleótido (polimorfismos de sequência) que quando se encontram em pelo menos 1% da população são denominados de SNP (Figura 1.7.1). Estes podem ser bi-, tri- ou tetra-alélicos, porém os tri- e tetra-alélicos são raros na população humana, pelo que se consideram estes marcadores como sendo bialélicos (Carvalho, 2010).

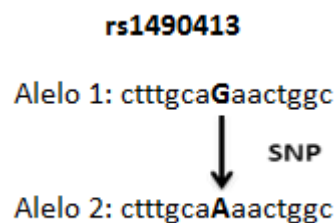


Figura 1.7.1 – Exemplo do SNP, identificado por rs1490413, onde ocorre substituição de uma Guanina por uma Adenina, podendo o alelo apresentar duas formas diferentes.

Os SNP são dos polimorfismos mais abundantes no genoma, pelo que podem ser utilizados para diferenciar indivíduos, facto este, que despertou o interesse dos especialistas em Genética Forense (Butler, 2010a). É possível encontrar um SNP em cada 200 a 300pb (Carvalho, 2010) e estima-se que existam cerca de 1 milhão de diferenças entre dois genomas, devidas aos SNP, o que representa aproximadamente 85% da variação genética em humanos (Goodwin, 2007b). Estima-se também que existam cerca de 3 milhões de SNP comuns numa população, com frequências superiores a 5% (Carracedo e Sanchez, 2005).

Estes marcadores podem ser classificados, de acordo com o seu uso, em SNP de identificação humana, SNP de ancestralidade, SNP de linhagem e SNP fenotípicos (Carvalho, 2010).

Os SNP utilizados em identificação humana, tal como outros tipos de marcadores, encontram-se em regiões autossómicas e não codificantes do genoma e apresentam, como todos os marcadores utilizados em Genética Forense, elevados níveis de polimorfismo que permitem a distinção de indivíduos.

Quanto aos SNP de ancestralidade, também denominados de AIM, estes permitem inferir acerca da origem geográfica de um dado indivíduo, o que permite por sua vez ter uma ideia do aspeto físico desse mesmo indivíduo. Isto é possível porque estes marcadores apresentam frequências alélicas que variam de população para população, o que requer a existência de bases de dados de frequências alélicas para as diversas populações.

Os SNP de linhagem compreendem os SNP do cromossoma Y (linhagens paternas) e do ADNmt (linhagens maternas), que são de herança uniparental e que não sofrem recombinação genética. Como as mutações que ocorrem nestes locais são transmitidas de geração em geração, tornam-se assinaturas genéticas da história da evolução de um *locus*, sendo possível realizar estudos de evolução e análises de parentesco com este tipo de marcadores.

Por fim, os SNP fenotípicos, que representam uma área recente, baseiam-se na análise de SNP em genes envolvidos na determinação de características fenotípicas, como a cor dos olhos, a cor do cabelo e a pigmentação da pele, de forma a prever essas mesmas em amostras biológicas provenientes de investigações criminais e para identificação humana. Porém é de salientar que esta classe de SNP é

menos informativa do que os STRs correntemente utilizados na determinação de perfis genéticos (Carvalho, 2010; Goodwin *et al.*, 2007b).

1.7.1. Vantagens e Desvantagens

Os marcadores SNP apresentam diversas vantagens (na maioria relativas aos STRs) que os tornam bastante úteis em Genética Forense. Uma dessas vantagens é o tamanho reduzido dos amplicões gerados após a amplificação por PCR, que apresentam menos de 100pb. Isto torna-se importante pois permite a análise de ADN degradado ou em pequenas quantidades. Também pelo facto de apresentarem taxas de mutação entre 10^{-8} e 10^{-9} , menores que os 10^{-4} dos STRs (Brumfield *et al.*, 2003), os SNP são mais vantajosos para investigações de parentesco aquando da ocorrência de pequenas incompatibilidades e para investigações de casos de imigração. Este facto também permite o uso dos SNP como AIM, pois existe uma maior probabilidade de se fixarem numa determinada população. Outra vantagem dos SNP está relacionada com o processamento e a análise, na medida em que estes marcadores são passíveis de serem analisados em primeiro lugar, através de reações de PCR *multiplex* com maior número de marcadores do que as existentes para os STRs, e em segundo lugar com recurso a tecnologias de elevado rendimento e automatizadas, já utilizadas para as investigações com STRs. Ainda relacionado com a análise, quando esta se processa por eletroforese capilar, a inexistência de artefactos, simplifica a análise dos perfis genéticos obtidos, novamente comparativamente aos STRs. Por fim, como estes são bialélicos, o estudo e a validação destes é mais simples, com recurso a menos amostras para o cálculo das frequências de cada alelo numa população (Butler, 2010a; Carvalho, 2010; Phillips, 2005; Sanchez *et al.*, 2006; Sobrino, *et al.*, 2005).

Apesar de todas estas vantagens, estes marcadores não são muito utilizados em investigações forenses, pois também apresentam diversas desvantagens que limitam a sua utilização. Em primeiro lugar, para obter uma probabilidade de *matching* e um poder de discriminação similar aos dos STRs, é necessário o estudo de um maior número de marcadores. Diferentes autores sugerem um diferente número de marcadores necessários, mas em média será necessário estudar entre 50 e 100 marcadores para obter o mesmo poder que 10 a 16 marcadores STR. Este número elevado de marcadores a serem analisados em simultâneo requer uma seleção de múltiplos SNP por cromossoma, o que requer por sua vez que sejam

realizados estudos para verificar se existe LD entre os marcadores. Outra desvantagem prende-se com o facto de serem bialélicos, o que dificulta a interpretação de perfis genéticos que apresentem misturas, pois é difícil a distinção entre um *locus* verdadeiramente heterozigótico de um homozigótico para ambos os dadores da mistura. Também por serem bialélicos, estes marcadores são menos informativos que os STRs que apresentam múltiplos alelos para cada *locus* (Butler, 2010a; Carvalho, 2010; Phillips, 2005; Sanchez *et al.*, 2006; Sobrino *et al.*, 2005). As principais diferenças entre os marcadores SNP e STR encontram-se na Tabela 1.7.1.

Tabela 1.7.1 – STRs vs SNP. Adaptado de Carvalho, 2010

Características	STR	SNP
Poder informativo	Elevado	Baixo
Tipo de marcador	di-,tri-, tetra- ou penta-nucleotídico	Bi-alélico
Nº alelos por marcador	Mais de 5	2
Maior vantagem em Genética Forense	Facilidade de interpretação de misturas	Facilidade de análise de amostras degradadas

1.7.2. Métodos de Tipagem

Não existe um método ideal para a tipagem de SNP, existem sim diversas metodologias cuja escolha depende da aplicação que se pretende. Na área forense, deve ter-se em atenção a escolha de um método de elevado rendimento, simples e de elevada reprodutibilidade.

A tipagem de SNP requer a deteção do(s) nucleótido(s) que constitui(em) cada SNP nos diferentes *loci*. Os diversos métodos de tipagem podem ser divididos em quatro grupos que diferem no mecanismo molecular em que se baseiam: hibridação específica de alelos, extensão de *primers*, ligação de oligonucleótidos e clivagem invasiva. Cada um destes pode ser detetado por diversas metodologias, como por exemplo por fluorescência (Sobrino *et al.*, 2005).

Na área forense, uma das metodologias mais utilizadas envolve a minisequenciação (Kuppuswamy *et al.*, 1991; Sokolov, 1990; Syvänen *et al.*, 1990) (Figura 1.7.2), também denominada de reação de extensão de *primers*, que é uma versão mais simples da sequenciação de ADN desenvolvida por Sanger em 1977, e que permite a análise simultânea de marcadores que se encontram distribuídos por todo o genoma (Goodwin *et al.*, 2007b). Este método envolve numa primeira fase a

amplificação das regiões alvo por PCR. A reação de extensão que se procede de seguida consiste na utilização de *primers* que se ligam ao produto da PCR e cuja extremidade 3' se liga ao local imediatamente adjacente ao SNP. A esses *primers* é adicionado (fase de extensão) um dideoxinucleotido (ddNTP) marcado com fluorescência, complementar ao SNP presente no ADN molde. A análise dos produtos desta reação é realizada por eletroforese capilar com detecção por fluorescência, que é uma tecnologia presente na maioria dos laboratórios de Genética Forense (Goodwin *et al.*, 2007b; Sobrino *et al.*, 2005).

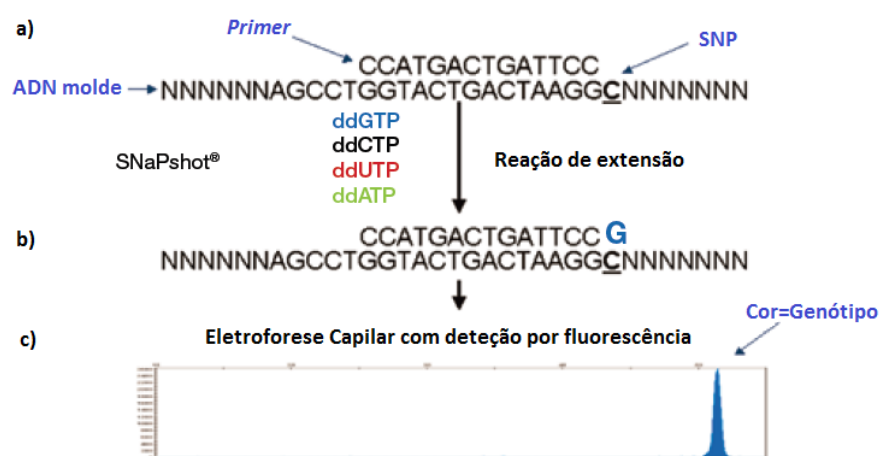


Figura 1.7.2 – Reação de Minisequencição. a) Os produtos da PCR são utilizados como ADN molde na reação de extensão; b) são utilizados *primers* que se ligam ao local adjacente ao polimorfismo e um ddNTP marcado com fluorescência complementar ao SNP é adicionado; c) o SNP é detetado em forma de pico por eletroforese capilar com detecção por fluorescência. Adaptado de Biosystems, 2012

1.7.3. Bases de Dados e Painéis de SNP

Uma das vantagens da minisequencição é permitir a análise simultânea de diferentes marcadores. Pelo facto de existirem SNP que apresentam frequências alélicas semelhantes e outros que apresentam frequências alélicas diferentes nas principais populações, é possível criar um painel de marcadores que permita a identificação humana na maioria das populações e complementar esse painel com marcadores específicos para uma determinada população.

Em 2003 foi criado o consórcio SNPforID (<http://spsmart.cesga.es/snpforid.php>). Este consórcio tinha como objetivos a seleção de pelo menos 50 SNP

autossômicos para identificação humana e as respectivas frequências alélicas nos maiores grupos populacionais, o desenvolvimento e avaliação de métodos de amplificação e detecção desses marcadores e a avaliação da aplicação dos mesmos na área forense (Amigo *et al.*, 2008). Com base nestes objetivos, Sanchez e colaboradores (2006), desenvolveram e avaliaram um painel de 52 marcadores SNP autossômicos (52-plex), polimórficos nas três principais populações (Europeia, Asiática e Africana) e que pode ser estudado com recurso às tecnologias presentes nos laboratórios de Genética Forense. Os critérios para a seleção dos marcadores foram: tamanho dos amplicões gerados inferior a 120pb, heterozigocidade mínima de 30% em pelo menos uma das populações e de 20% nas três populações, localizados nos braços curto e longo de cada autossoma com uma distância mínima de 100kb entre os SNP candidatos e os genes vizinhos, sem associação com os STRs utilizados na rotina forense e com sequências flanqueadoras já descritas, sem polimorfismos que possam interferir com a ligação dos *primers*. Outro painel, de 34 marcadores SNP informativos de ancestralidade (34-plex) foi também desenvolvido e incluído na base de dados SNPforID, para estudos de ancestralidade (Phillips *et al.*, 2007).

Atualmente, o SNPforID é uma base de dados pública que contém informação acerca das frequências alélicas para cada SNP pertencente tanto ao 52-plex como ao 34-plex nas variadas populações. Permite também a comparação de dados com outras bases de dados existentes (1000 Genomes, HapMap, Perlegen e HGDP-CEPH, disponíveis em <http://spsmart.cesga.es/>) e a comparação de dados entre as diversas populações presentes nas mesmas. Estas bases de dados encontram-se em constante atualização e expansão, conforme são adicionados novos dados e novas populações (Amigo *et al.*, 2008). À data da última atualização (Novembro 2013), o SNPforID apresenta dados relativamente a 3 593 indivíduos de 84 origens diferentes para o 52-plex e dados acerca de 2 018 indivíduos de 69 origens para o 34-plex. Foi a partir desta(s) base(s) de dados que outros painéis de SNP foram desenvolvidos e validados, para identificação humana e para estudos de ancestralidade (Børsting *et al.*, 2009; Bulbul *et al.*, 2009; Dixon *et al.*, 2005; Fang *et al.*, 2009; Freire-Aradas *et al.*, 2012; Kosoy *et al.*, 2009; Nievergelt *et al.*, 2013; Phillips *et al.*, 2007; Sanchez *et al.*, 2008; Whittle *et al.*, 2009).

Para além do SNPforID, existe outra base de dados pública, ALFRED (do inglês *allele frequency database*), que contém informação acerca das frequências

alélicas dos marcadores SNP em diversas populações (disponível em <http://alfred.med.yale.edu/>). Nesta estão incluídas dados do SNPforID e também serviu como base ao desenvolvimento de painéis de SNP para identificação e para estudos de ancestralidade (Kidd *et al.*, 2006, 2012). Esta base de dados inclui não só dados relacionados com os SNP mas também com outros tipos de polimorfismos existentes no genoma humano.

Existe ainda outra base de dados importante, a *dbSNP* (disponível em www.ncbi.nlm.nih.gov/snp) onde existem desde a última atualização (atualização 137), cerca de 61 milhões de SNP descritos e identificados (em regiões codificantes e não codificantes do genoma) e, de onde se pode retirar toda a informação acerca dos mesmos. A partir desta base de dados foram também desenvolvidos painéis de marcadores SNP para fins forenses (Vallone 2005, 2006).

Foi com base no painel de 52 marcadores SNP, que Dario e colaboradores (2009) desenvolveram um painel de 20 marcadores SNP, em continuação do trabalho realizado por Costa e colaboradores (2008), passíveis de serem analisados em duas reações de PCR *multiplex*, com 10 marcadores cada, como um método alternativo aos STRs em investigações de paternidade complexas. Os marcadores foram selecionados por apresentarem maiores níveis de polimorfismo (aproximadamente 50% de heterozigotia) na população portuguesa, espanhola (mais concretamente Galiza, pela proximidade a Portugal) e africana. As informações relativas a esses marcadores encontram-se na Tabela 1.7.2 para o Auto 3 (denominação dada ao *multiplex*) e na Tabela 1.7.3 para o Auto 4 (denominação dada ao *multiplex*).

Tabela 1.7.2 – Marcadores SNP pertencentes ao PCR *multiplex* Auto 3: respectivas identificações na base de dados do NCBI, localizações genómicas e cromossómicas, identificação segundo o consórcio SNP (TSC), variação nucleotídica e tamanho de amplificação gerado em pb. (SI – sem informação)

SNPforID	ID NCBI <i>dbSNP</i>	Cromossoma	Posição (nt)	TSC#	SNP	Tamanho amplificação (pb)
A01	rs1490413	1	4 037 521	TSC0724193	G-A	68
A06	rs1029047	6	1 080 939	TSC0253802	T-A	100
A08	rs763869	8	1 363 017	TSC0065968	C-T	100
A10	rs735155	10	3 328 178	TSC0027519	G-A	100
A12	rs2107612	12	741 262	TSC1108144	G-A	93
A14	rs1454361	14	23 840 960	TSC0684657	A-T	73
A27	rs2111980	12	104 830 721	TSC1113476	A-G	72
A49	rs1005533	20	40 172 539	TSC0082071	G-A	107
A50	rs8037429	15	51 332 965	SI	C-T	108
A51	rs891700	1	236 923 075	TSC0162577	A-G	109

Tabela 1.7.3 – Marcadores SNP pertencentes ao PCR *multiplex* Auto 4: respectivas identificações na base de dados do NCBI, localizações genómicas e cromossómicas, identificação segundo consórcio SNP (TSC), variação nucleotídica e tamanho de amplificação gerado em pb.

SNPforID	ID NCBI <i>dbSNP</i>	Cromossoma	Posição (nt)	TSC#	SNP	Tamanho amplificação (pb)
A04	rs2046361	4	10 719 942	TSC1065282	A-T	79
A05	rs717302	5	2 932 133	TSC0039610	G-A	86
A13	rs1886510	13	20 172 700	TSC0904551	C-T	86
A16	rs729172	16	5 606 490	TSC0028090	C-A	60
A29	rs1024116	18	73 559 363	TSC0247167	G-A	76
A35	rs1463729	9	122 257 493	TSC0377760	G-A	87
A36	rs2076848	11	134 205 198	TSC0022275	T-A	89
A37	rs1355366	3	192 127 021	TSC0491662	A-G	90
A38	rs907100	2	239 850 329	TSC0186810	G-C	91
A41	rs737681	7	154 850 085	TSC0033074	T-C	96

Como todos os painéis de marcadores genéticos utilizados em análises forenses, estes 20 marcadores encontram-se distribuídos por todo o genoma, com exceção dos cromossomas mais pequenos. Na Figura 1.7.3 pode-se observar a distribuição cromossómica dos 20 SNP.

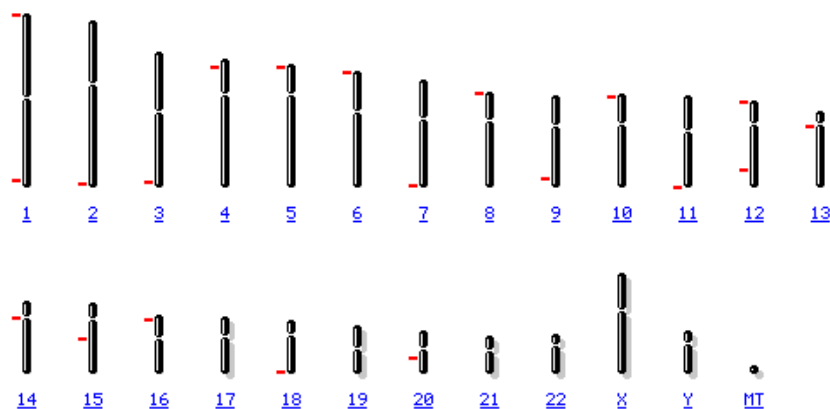


Figura 1.7.3 – Distribuição dos 20 marcadores SNP ao longo dos diferentes cromossomos. Cada cromossoma está identificado pelo respetivo número em baixo (a azul) e os traços vermelhos representam a localização de cada SNP nos diversos cromossomas. Imagem criada a partir de <http://www.ncbi.nlm.nih.gov/> [Consultado em 14.12.2013]

1.8. Marcadores do tipo InDel

Mais recentemente, outro tipo de marcadores genéticos (InDel) têm vindo a ganhar interesse quer ao nível da Genética Clínica quer ao nível da Genética Forense.

Os marcadores do tipo InDel são polimorfismos de inserção/delecção, também denominados de DIP (do inglês *deletion/insertion polymorphism*). Estes são polimorfismos de comprimento, originários de mutações de inserção ou deleção, ou seja, são caracterizados pela presença ou ausência de uma sequência específica de nucleótidos no ADN (Fondevila *et al.*, 2011a). Um exemplo de um polimorfismo InDel está representado na Figura 1.8.1.

rs16363

Inserção: tgaatcc**TGTTT**tgttttg

Deleção: tgaatcc_____tgttttg

Figura 1.8.1 – Exemplo de um InDel, rs16363, onde se pode observar a variante com a inserção de 5 nucleótidos TGTTT e a variante com a deleção desses mesmos 5 nucleótidos, com alteração do tamanho da sequência.

Na última década foram desenvolvidos diversos trabalhos com o intuito de identificar as variantes InDel existentes no genoma humano e de entender de que forma estas afetam a funcionalidade dos genes (quer ocorram em regiões codificantes ou em regiões não codificantes do genoma). Os primeiros estudos de identificação de InDel focaram-se na análise do cromossoma 22 (Dawson, 2001; Mullikin *et al.*, 2000), onde, pelos resultados obtidos, se previu uma grande abundância de variantes do tipo InDel no genoma humano. Em 2002, Weber e colaboradores, identificaram 2 mil variantes de inserção/deleção, distribuídas por todo o genoma, por comparação de sequências genômicas e de ADN complementar. A maioria destas variantes são pequenas inserções/deleções com um tamanho entre 2 a 16pb, com um comprimento máximo de 55pb. Três anos depois, Bhangale e colaboradores (2005), utilizaram a técnica de sequenciação com detecção por fluorescência, já utilizada para a detecção de SNP, de forma a ultrapassar algumas falhas relativas à comparação genômica que não permitia a detecção de polimorfismos heterozigóticos. Estes autores detetaram 2 393 InDel em 330 genes que apresentavam entre 1 e 543pb, sendo que a maioria apresentava um comprimento menor que 5pb. Verificam também, tal como Weber (2002), que as frequências destas variantes diminuem com o aumento do comprimento do InDel. Já em 2006, Mills e colaboradores, desenvolveram um mapa inicial das variantes do tipo InDel que continha 415 436 variantes identificadas em três populações humanas e que apresentavam um comprimento entre 1pb e 9 989pb. Estes autores também verificaram que os InDel apresentam uma densidade média de 1 por cada 7,2kb de ADN, variando esse número conforme a quantidade de ADN presente em cada cromossoma. Chegaram também à conclusão de que a população humana deve conter cerca de 1,56 milhões de InDel. Kidd e colaboradores (2008) identificaram quase 800 mil InDel, dos quais 525 não estavam publicados e em 2011, Mills e colaboradores identificaram cerca de 2 milhões de pequenos InDel no genoma humano, dos quais apenas 800 mil se encontravam nas regiões codificantes do genoma. A maioria destes InDel apresentavam um tamanho menor que 100pb e 63% destes eram novos, em relação aos já publicados anteriormente na base de dados *dbSNP* (atualização 129) (<http://www.ncbi.nlm.nih.gov/SNP/>). Todos estes autores verificaram que as variações de inserção/deleção são, depois dos SNP, as mais abundantes no genoma humano e que são úteis como marcadores moleculares tanto

ao nível da Genética Clínica como da Genética Forense e da Genética Populacional/Evolutiva. Para além destes, outros estudos foram realizados em genomas/populações específicos (Ahn *et al.*, 2009; Bentley *et al.*, 2008; Kim *et al.*, 2009; Levy *et al.*, 2007; Ley *et al.*, 2008; Wang *et al.*, 2008; Wheeler *et al.*, 2008). Na Tabela 1.8.1 pode-se observar um resumo de alguns estudos publicados acerca da identificação das variantes do tipo InDel referidas.

Atualmente, existem na base de dados *dbSNP* (atualização 137) (www.ncbi.nlm.nih.gov/snp) 11 milhões de InDel descritos ao longo de todo o genoma humano que representam cerca de 18 a 25% de todos os polimorfismos existentes.

Os InDel podem ser divididos em 5 classes: inserções e deleções de um par de bases; expansões monoméricas; expansões de unidades de repetição com 2 a 15pb; inserções de transposões e sequências aleatórias de ADN (Mills *et al.*, 2006).

Tabela 1.8.1 - Estudos publicados sobre variantes InDel descobertas no genoma humano, os respetivos números de InDel e tamanhos. (SI – sem informação). Adaptado de: Mullaney e Mills, 2010

Estudo	Nº InDel	Tamanho InDel (pb)
Mullikin (2000)	NR	SI
Dawson (2001)	2 180	SI
Weber (2002)	2 000	2-55
Bhangale (2005)	2 393	1-543
Mills (2006)	415 436	1-9989
Kidd (2008)	796 273	1-100
Mills (2011)	1,96 Milhões	1-10 000
Levy (2007)	823 396	1-82 711
Wheeler (2008)	222 718	2-38 896
Wang (2008)	135 262	1-3
Bentley (2008)	400 000	1-16
Ley (2008)	726	1-30
Ahn (2009)	342 965	1-26
Kim (2009)	170 202	1-29

1.8.1. Vantagens e Desvantagens

As vantagens que os marcadores do tipo InDel apresentam são a principal razão pela qual têm vindo a ganhar interesse nos últimos anos. Essas vantagens são uma combinação das características vantajosas dos marcadores SNP e STR (uma breve comparação entre os marcadores STR e InDel está apresentada na Tabela 1.8.2).

Em primeiro lugar, os InDel encontram-se distribuídos por todo o genoma, isto é, em regiões codificantes e não codificantes e em cromossomas autossómicos e sexuais com mais ou menos a mesma densidade (apesar de existirem *hotspots*), o que permite um grande variedade de combinações de marcadores na construção de painéis de estudo, consoante a aplicação que se pretende. Em segundo lugar, têm origem num único evento de mutação, estável e cujas frequências são baixas (comparativamente aos STRs). Em terceiro lugar, devido ao seu pequeno tamanho, dão origem a fragmentos também pequenos, o que os torna vantajosos na análise de ADN degradado (Romanini *et al.*, 2012) e também o que permite serem estudados em grandes *multiplex*. Outra vantagem é o facto de apresentarem diferenças nas frequências alélicas entre populações afastadas geograficamente, podendo serem utilizados como marcadores AIM. Por fim, a genotipagem é relativamente simples e rápida (comparativamente aos SNP), requerendo métodos pouco laboriosos e passíveis de serem automatizados (Costa Francez *et al.*, 2012; Fondevila *et al.*, 2012; Pereira *et al.*, 2009a).

Tabela 1.8.2 – STRs vs InDel. Adaptado de: Butler, 2010b

Características	STR	InDel
Frequência no genoma humano	~1 em cada 15kb	~1 em cada 7.2kb
Poder informativo	Elevado	Baixo
Tipo de marcador	di-,tri-, tetra- ou penta-nucleotídico	Bi-alélico
Nº alelos por marcador	Mais de 5	2
Tamanho dos amplicões	~75-400pb	<160pb

Quanto às desvantagens, destaca-se o facto de serem marcadores bialélicos, o que dificulta a análise de perfis genéticos com misturas de dois ou mais dadores. Esta dificuldade é devida à difícil distinção entre um alelo heterozigótico (que é na realidade homozigótico em ambos os dadores) e um verdadeiro heterozigótico para um dos dadores. Também por apresentarem um poder de discriminação inferior aos

STRs, é necessário o estudo de um maior número de marcadores, o que implica uma análise mais cuidadosa e o desenvolvimento de reações *multiplex* maiores e mais robustas.

1.8.2. Métodos de Análise

Com a identificação dos polimorfismos de inserção/deleção, veio a necessidade de desenvolver técnicas de análise que permitam uma rápida (e simples) genotipagem destes marcadores autossômicos.

Em 2009, Pereira e colaboradores, desenvolveram um PCR *multiplex* que permite a análise de um painel de 38 marcadores InDel (também este desenvolvido pelos autores), com base no trabalho desenvolvido por Weber e colaboradores (2002). A seleção dos marcadores, dentro dos cerca de 4 mil InDel disponíveis à data, obedeceu a critérios específicos, entre os quais, serem bialélicos e não codificantes, apresentarem frequências alélicas mínimas maiores que 0,25 nas três principais populações (Europeia, Africana e Asiática), apresentarem uma heterozigocidade média maior que 0,40 e terem uma variação de tamanho dos alelos entre 2 a 5pb. Outros painéis de marcadores têm vindo a ser desenvolvidos com o intuito de serem aplicados em investigações de parentesco (Pimenta e Pena, 2010).

Para além deste painel de 38 marcadores, existe outro, comercialmente disponível, denominado de Investigator[®]DIPplex e desenvolvido pela *QIAGEN*. Este *kit* permite a análise simultânea de 30 marcadores InDel autossômicos e não codificantes e a análise da Amelogenina para os cromossomas sexuais. Tal como o painel desenvolvido por Pereira e colaboradores (2009), estes marcadores encontram-se marcados com 4 fluorocromos que emitem fluorescência em comprimentos de onda diferentes, facilitando a análise. Os marcadores incluídos neste painel, encontram-se descritos na Tabela 1.8.3. Estes encontram-se distribuídos por 19 cromossomas autossômicos, separados por pelo menos 10Mpb (mega pares de bases) dos marcadores STR e SNP comercialmente disponíveis (o que permite o seu uso em complementaridade com esse marcadores) e apresentam um tamanho variável entre 4 e 22pb. Outra característica destes marcadores é o tamanho dos fragmentos gerados, que apresentam um tamanho máximo de 160pb (*QIAGEN*, 2013).

Tabela 1.8.3 – Painel de 30 marcadores InDel e a Amelogenina incluídos no *kit* de Investigator®DIPplex (QIAGEN). Pode-se observar o nome dado ao marcador, respectiva referência na base de dados *dbSNP*, localização cromossômica e sequência de inserção/deleção (HLD do inglês *Human Locus DIP*). Adaptado de: QIAGEN, 2013.

DIP locus	ID <i>dbSNP</i>	Localização	Sequência
Amelogenina X	M55418	Xp22.1-22.3	-
Amelogenina Y	M55419	Yp11.2	-
HLD77	rs1611048	7q31.1	TAAG
HLD45	rs2307959	2q31.1	CACG
HLD131	rs1611001	7q36.2	TGGGCTTATT
HLD70	rs2307652	6q16.1	AGCA
HLD6	rs1610905	16q13	GCAGGACTGGGCACC
HLD111	rs1305047	17p11.2	CACA
HLD58	rs1610937	5q14.1	AGGA
HLD56	rs2308292	4q25	TAAGT
HLD118	rs16438	20p11.1	CCCCA
HLD92	rs17174476	11q22.2	GTTT
HLD93	rs2307570	12q22	ACTTT
HLD99	rs2308163	14q23.1	TGAT
HLD88	rs8190570	9q22.32	CCACAAAGA
HLD101	rs2307433	15q26.1	GTAG
HLD67	rs1305056	5q33.2	CTACTGAC
HLD83	rs2308072	8p22	AAGG
HLD114	rs2307581	17p13.3	TCCTATTCTACTCTGAAT
HLD48	rs28369942	2q11.2	GACTT
HLD124	rs6481	22q12.3	GTGGA
HLD122	rs8178524	21q22.11	GAAGTCTGAGG
HLD125	rs16388	22q11.23	ATTGCC
HLD64	rs1610935	5q12.3	GACAAA
HLD81	rs17879936	7q21.3	GTAAGCATTGT
HLD136	rs16363	22q13.1	TGTTT
HLD133	rs2067235	3p22.1	CAACCTGGATT
HLD97	rs17238892	13q12.3	AGAGAAAGCTGAAG
HLD40	rs2307956	1p32.3	GGGACAGGTGGCCACTAGGA GA
HLD128	rs2307924	1q31.3	ATTAAATA
HLD39	rs17878444	1p22.1	CCTAAACAAAAATGGGAT
HLD84	rs3081400	8q24.12	CTTTC

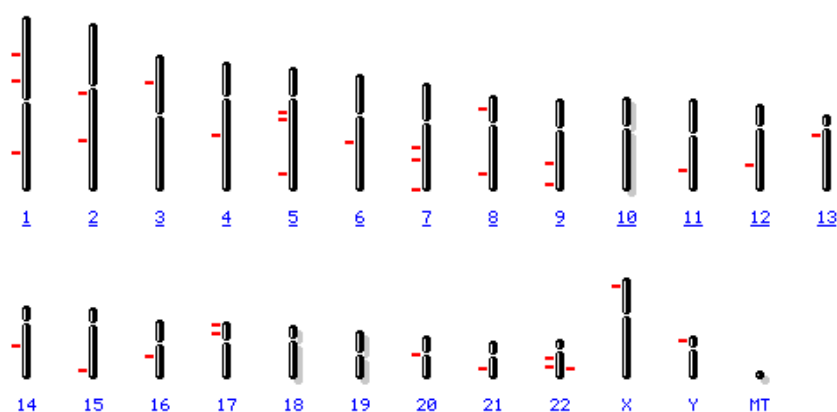


Figura 1.8.2 – Distribuição dos 30 marcadores InDel incluídos no *kit* de Investigator[®]DIPplex (QIAGEN) ao longo dos diferentes cromossomas. Cada cromossoma está identificado pelo respetivo número em baixo (a azul) e os traços vermelhos representam a localização de cada InDel nos diferentes cromossomas. Imagem criada a partir de <http://www.ncbi.nlm.nih.gov> [Consultado em 14.12.2013]

Ambos os painéis (Investigator[®]DIPplex e 38-plex) permitem a análise de pequenas quantidades de ADN, pelo facto de gerarem amplicões de pequeno tamanho e foram desenvolvidos com o objetivo de identificação humana. Em conjunto, ou seja, o uso de cerca de 60 marcadores InDel em dois *multiplex* simples apresenta uma vantagem quando se obtém perfis de STR incompletos, porém o *kit* comercialmente disponível permite a obtenção de perfis mais balanceados que o 38-plex, o que o torna mais vantajoso na análise de misturas (Fondevila *et al.*, 2012).

Estes conjuntos de marcadores foram construídos para serem analisados por eletroforese capilar com deteção por fluorescência, com recurso a um sequenciador automático presente em praticamente todos os laboratórios de Genética Forense. A pensar nos países onde os serviços forenses são escassos e/ou onde não existem fundos para obter as tecnologias utilizadas na Genética Forense, Pena e Pena (2012) desenvolveram um painel de com 40 marcadores InDel que pode ser analisado em gel de poliacrilamida com coloração com nitrato de prata, com a grande vantagem de ser mais económico.

Um dos potenciais usos dos polimorfismos InDel é como marcadores de ancestralidade, devido às diferenças das frequências alélicas entre as populações da Europa, África e Ásia e tal como foram desenvolvidos painéis de marcadores do tipo InDel para identificação individual, também foram desenvolvidos painéis que

permitem inferir acerca da ancestralidade geográfica e que permitem perceber as relações genéticas em populações que sofreram diversas misturas. Um painel de 48 marcadores foi desenvolvido por Santos e colaboradores (2010), com base na população Brasileira, que permite distinguir populações pelo continente de origem e também as subpopulações que se originaram aquando da mistura. Também Pereira e colaboradores (2012) desenvolveram um painel de 46 marcadores de ancestralidade InDel que pode ser utilizado para atribuir uma população de origem a um indivíduo em casos forenses, tal como o painel de 21 marcadores desenvolvido por Zaumsegl e colaboradores (2013).

1.9. Objetivos

Tendo em conta o número elevado de imigrantes oriundos da República da Guiné-Bissau e da República de Cabo Verde a residir legalmente em Portugal, tal como as ligações históricas entre estes países e entre Portugal, este estudo tem como objetivo a realização da caracterização genética das populações cabo-verdiana e guineense, imigrantes em Portugal, mais especificamente na região de Lisboa. Para além disso, pretende-se utilizar, para essa caracterização, marcadores genéticos forenses recentes, InDel e SNP, que apresentam características vantajosas relativamente aos STRs muito utilizados. De uma forma geral, pretende-se comparar as duas populações entre si e com a população onde estão atualmente inseridos, e de modo a ser possível a utilização da informação obtida na rotina forense e criminal de identificação genética, no INMLCF-Sul.

Mais especificamente pretende-se:

- Estudar indivíduos das populações da Guiné e de Cabo Verde com PCR *multiplex* de marcadores do tipo InDel com capacidade para definir perfis genéticos individuais;
- Estudar esses mesmos indivíduos com PCR *multiplex* de marcadores do tipo SNP, de forma a complementar os perfis genéticos de InDel;
- Calcular as frequências alélicas dos marcadores estudados nas populações de Guiné e Cabo Verde;
- Calcular outros parâmetros populacionais e forenses dos marcadores estudados nas populações de Guiné e Cabo Verde, designadamente equilíbrio de Hardy-Weinberg (HWE), heterozigotia observada (H_o),

heterozigotia esperada (He), probabilidade de *matching* (MP), poder de discriminação (PD), *polymorphism information content* (PIC), poder de exclusão (PE) e índice de paternidade associado (TPI);

- Comparação dos parâmetros forenses e populacionais obtidos nas populações estudadas com a população de Portugal e com outras populações tipadas com os mesmos marcadores genéticos.

2. Materiais e Métodos

2.1. Procedimento Experimental

A realização do trabalho requereu diversas fases laboratoriais para cada um dos tipos de marcadores utilizados, estando identificadas essas fases no fluxograma a baixo apresentado.

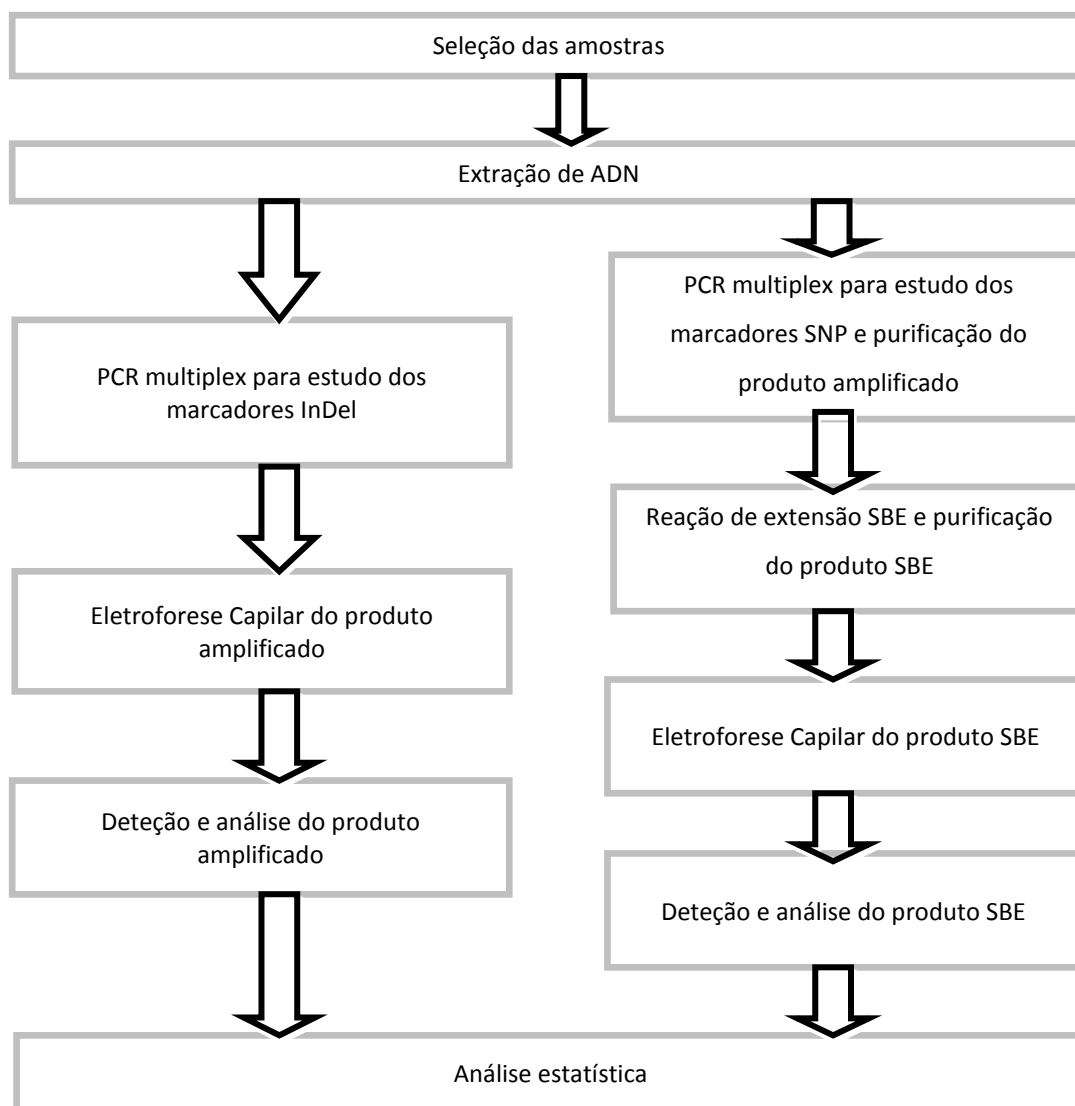


Figura 2.1.1 – Fluxograma representativo das diversas fases laboratoriais para a realização do estudo dos marcadores InDel e dos marcadores SNP.

2.2. Seleção das Amostras

Neste trabalho foram estudados um total de 450 indivíduos aparentemente saudáveis, de origem cabo-verdiana ou guineense, residentes na região Sul de Portugal, os quais compareceram ao Serviço de Genética e Biologia Forenses (SGBF) da delegação do Sul do INMLCF, no âmbito de ações de investigação de parentesco biológico.

Do total de indivíduos estudados, 367 têm origem paterna e materna cabo-verdiana e 83 têm origem paterna e materna guineense.

As amostras de sangue selecionadas foram colhidas no SGBF entre os anos de 2002 e 2012.

2.3. Extração de ADN

A extração de ADN das amostras foi realizada pelo método Chelex 100[®], (BioRad), descrito em 1991 por Walsh e seus colaboradores (Walsh *et al.*, 1991). O Chelex é uma resina quelante que apresenta uma elevada afinidade para iões metálicos polivalentes, o que permite a prevenção da degradação do ADN a elevadas temperaturas.

Começou-se por retirar 1 círculo, do cartão *Whatman™ FTA™* (GE Healthcare) onde havia sido colhida a amostra, com um *paper punch single hole* e colocou-se em tubos *eppendorf*, previamente identificados. Aos tubos, adicionou-se 1mL de água Milli-Q e deixou-se a incubar à temperatura ambiente, durante cerca de 15 minutos. Agitaram-se os tubos no *vortex*, a uma baixa velocidade, durante 5 a 10 segundos, centrifugou-se a 14 000rpm durante 3 minutos a uma temperatura de 20°C e retirou-se cerca de 960μL de sobrenadante de cada amostra.

Ao *pellet* foi adicionado 180μL de Chelex a 5%, previamente preparado, seguido de um banho de aquecimento a 56°C durante 15 minutos. Procedeu-se a uma nova agitação, a alta velocidade, durante 10 a 15 segundos, seguida de um banho em ebulição durante 8 minutos. Realizou-se nova agitação a alta velocidade durante alguns segundos e nova centrifugação a 14 000rpm durante 5 minutos.

O ADN extraído foi guardado a -20°C até ser necessário para a amplificação.

2.4. Estudo dos marcadores InDel

Para o estudo dos marcadores InDel foi utilizado o *kit* comercial, Investigator®DIPplex (QIAGEN), o qual permite analisar 30 marcadores InDel autossómicos e o gene da Amelogenina para a identificação do género.

O *kit* em uso contém todos os reagentes necessários ao estudo dos 30 marcadores InDel, como se pode observar na Tabela 2.4.1. Dos constituintes deste *kit* destaca-se a mistura de *primers* que permite a análise simultânea de todos os marcadores pelo facto destes *primers* marcados com fluorocromos que emitem fluorescência em quatro comprimentos de onda diferentes: no azul, no verde, no amarelo e no vermelho.

Tabela 2.4.1 – Constituintes do *kit* Investigator®DIPplex (QIAGEN)

Investigator®DIPplex	
Mistura de reação A	Marcador interno 550 (BTO)
Mistura de <i>primers</i> DIPplex	<i>Ladder</i> alélico DIPplex
ADN polimerase Multi Taq2	Água livre de nucleases
ADN controlo 9948	

Os marcadores InDel encontram-se distribuídos por 19 cromossomas e estão separados por pelo menos 10Mpb de cada marcador comercial utilizado na rotina da genética forense, STR e SNP (ver Tabela 1.8.3, página 38) (QIAGEN, 2013).

2.4.1. Amplificação por PCR *multiplex*

A amplificação dos 30 marcadores InDel e da Amelogenina decorreu numa única reação de PCR, denominada de PCR *multiplex*, onde é possível amplificar diferentes regiões do ADN numa única reação.

Para a reação de PCR *multiplex*, foi necessário preparar em primeiro lugar uma mistura de reação que contenha todos os componentes (incluídos no *kit* em uso) para que ocorra a amplificação dos marcadores em estudo. A mistura de reação é preparada para um volume de 10,5µL por amostra, aos quais se adiciona 2µL de solução com ADN alvo, perfazendo um volume final de reação de 12,5µL (da Silva *et al.*, 2013). Para os controlos de reação, positivo e negativo, foi colocado, em vez do ADN alvo em estudo, ADN controlo 9947A (*Applied Biosystems*) para o controlo positivo e água livre de nucleases para o controlo negativo, nas mesmas quantidades.

Tabela 2.4.2 – PCR *multiplex*. Reagentes e quantidades respectivas, necessários na preparação da mistura de reação para a amplificação dos marcadores InDel.

Componente	Volume por reação
Mistura de reação A	2,5µl
Mistura de <i>primers</i> DIPplex	2,5µl
Multi Taq2 ADN Polimerase	0,3µl
Água livre de nucleases	5,2µl
ADN alvo	2µl
Volume Total	12,5µl

A mistura de reação A contém os dNTPs necessários para a formação das novas cadeias de ADN, contem também MgCl₂ que influencia a capacidade de ligação dos *primers* ao ADN e contem também BSA, utilizado para aumentar o rendimento da amplificação. A mistura de *primers* contém todos os *primers* necessários para a amplificação dos 30 marcadores InDel e Amelogenina, marcados com quatro fluorocromos diferentes: 6-FAM (azul), BTG (verde), BTY (amarelo) e BTR (vermelho) (QIAGEN, 2013).

A reação de amplificação decorreu num termociclador *GeneAmp PCR system 9700 (Applied Biosystems)*, segundo as condições da Tabela 2.4.3, recomendadas para amostras em condições normais com 0,2 a 0,5ng de ADN.

Após a reação de amplificação pode-se guardar o produto amplificado a 4°C durante um período de tempo reduzido, ou a -20°C por longos períodos de tempo, ou pode-se seguir diretamente para a eletroforese capilar.

Tabela 2.4.3 – Condições de temperatura e tempo para que ocorra cada uma das fases necessárias à amplificação dos marcadores InDel e Amelogenina no termociclador.

Temperatura	Tempo
94°C	4min
94°C	30s
61°C	120s
72°C	75s
68°C	60min
10°C	∞

2.4.2. Eletroforese Capilar do produto amplificado

O produto amplificado obtido é uma mistura dos fragmentos de ADN dos 30 marcadores em estudo, de diferentes pesos moleculares, sendo por isso necessário separar e detetar esses fragmentos por eletroforese capilar com marcação por fluorescência, com recurso ao sequenciador automático *Genetic Analyzer 3130 xl* (*Applied Biosystems*).

Antes de se proceder à eletroforese capilar em si, preparou-se as amostras em estudo numa placa de 96 poços, *MicroAmp[®] Optical* (*Applied Biosystems*). Foi preparada uma mistura para a eletroforese contendo formamida Hi-Di (*Applied Biosystems*) que mantém as cadeias de ADN desnaturadas, e contendo o marcador interno de peso molecular 550 (BTO) (*Qiagen*), para um volume final por poço de 9,2µL (M Fondevila *et al.*, 2012). Após a distribuição das alíquotas por cada poço da placa adicionou-se 1µL de produto amplificado ou 1µL de *ladder* alélico, este último para permitir a identificação dos alelos aquando da análise. Procedeu-se em seguida a uma desnaturação a 95°C durante 3 minutos, no termociclador *GeneAmp PCR system 9700* (*Applied Biosystems*), seguida de um rápido arrefecimento em placa de gelo por mais 3 minutos, para que ocorra a separação das cadeias de ADN.

As condições da eletroforese capilar foram as especificadas pelo fabricante do *kit*, utilizando o polímero POP4[®] (*Applied Biosystems*).

Tabela 2.4.4 – Reagentes necessários para a preparação do produto amplificado para a eletroforese capilar e os respetivos volumes por cada reação.

Reagente	Volume por reação
Formamida Hi-Di	8,9µL
Marcador 550 (BTO)	0,3µL
Volume Total	9,2µL

2.4.3. Detecção e análise do produto amplificado

A análise da eletroforese capilar foi realizada no *software GeneMapper* versão 3.2 (*Applied Biosystems*), de acordo com as indicações do manual (QIAGEN, 2013). O limiar de deteção foi definido para um valor de 100 RFUs para os fragmentos marcados com fluorescência azul (B) e com fluorescência verde (G),

50RFUs para os fragmentos marcados com fluorescência amarela (Y), vermelha (R) e cor de laranja (O). Em relações às condições de leitura foram definidas para um *start point* de 1 750pb e um *stop point* de 6 000pb.

Para a interpretação dos resultados, foi utilizado o *freeware DIPSorter*. (QIAGEN)

2.4.4. Alelos *off ladder* e *dropout* do alelo D97+

A confirmação e verificação de fenómenos detetados, como a presença de alelos *off ladder* e a ineficiente amplificação do alelo D97+, aquando a análise dos eletroferogramas obtidos, foi realizada por repetição da tipagem dos indivíduos onde foram estes foram detetados. Também para entender melhor a causa/efeito destes fenómenos, foram genotipados 10 filhos desses mesmos indivíduos. A paternidade fora confirmada de forma a garantir a ocorrência dos fenómenos na descendência.

2.5. Estudo dos marcadores SNP

Foram estudados 20 marcadores SNP, disponíveis no laboratório e previamente selecionados por Dario e colaboradores em 2009 a partir do SNPforID 52-plex para identificação humana. A seleção destes marcadores teve em atenção assegurar que fossem polimórficos em populações Africanas. Os marcadores SNP em estudo encontram-se na Tabela 1.7.2 e na Tabela 1.7.3, bem como as informações acerca dos mesmos, separados por dois 10-plex, denominados de Auto 3 e Auto 4, desenhados através do SNPforID browser (Dario *et al.*, 2009). A metodologia utilizada foi descrita por Dario (2010).

2.5.1. Amplificação por PCR *multiplex* e purificação do produto amplificado

Numa primeira fase, é necessário amplificar as regiões imediatamente adjacentes aos locais onde ocorrem as variações.

A amplificação dos 20 marcadores SNP decorreu em 2 reações de PCR *multiplex* diferentes: uma reação para os marcadores pertencentes ao Auto 3 e outra para os pertencentes ao Auto 4.

Os *primers* utilizados para a reação de amplificação, descritos por Sanchez e colaboradores (2006) encontram-se descritos na Tabela 2.5.1 para o Auto 3, e na Tabela 2.5.2 para o Auto 4.

Tabela 2.5.1 – Primers utilizados para a PCR multiplex dos marcadores correspondentes ao Auto 3.
Adaptado de: Sanchez *et al.*, 2006

SNP	Primer forward (5'-3')	Primer Reverse (5'-3')
A01	GTGTGGACTGGGCTGATGT	TTCTCACTAGTGCCCTGCTCTG
A06	CATAACGTGGATTTGTCAGCA	GGAATAAACTGAAGGCTAAAGAAAAG
A08	ATCAAGTGCTTTCTGTTGACATTTG	GGCTACTCCCTCATAATGTAATGC
A10	GGAGAAAACCGGAGAGCTG	GAGTGTCACCGAATTCAACG
A12	GAGCATTCTCTTCTGTAAAAATTGC	TGAGTACATTATTCAACTGTTTTGGAG
A14	AGGGAAATACACCCTGAGCTG	AGCTGTCCATCATCAGTAAGACAC
A27	AGCATCTTGGCAGCATCC	AGCAAGATCTTTGCCAGTGAGT
A49	GGTTTGTGTGTGAGTGTTTCAGAT	CCTTATGCCTCCCCTGAAC
A50	TTCACTTTGCTACACCTCCATAGTA	TGCTACGTAAGAGGTCATTGCTATC
A51	TTTTCAGAGGTGGTATTCTAGCTG	GCTATGACACTCCTTAGAACTATGCAA

Tabela 2.5.2 – Primers utilizados para a PCR multiplex dos marcadores correspondentes ao Auto 4.
Adaptado de: Sanchez *et al.*, 2006

SNP	Primer forward (5'-3')	Primer Reverse (5'-3')
A04	CCTATTTGTATGTATCTATTGTCTATGAACG	GTCATTGTTGACACTTCACCTTCTA
A05	CTTTAGAAAGGCATATCGTATTAAGTGTG	AACACAGAAAGAGGTTTCATATGTTGG
A13	GTCCTTGTCAATCTTTCTACCAGAG	GGATTTTCACAACAACACTTGC
A16	CATTAATATGACCAAGGCTCCTCT	ACATTTCCCTCTTGCGGTTAC
A29	CCATGTGTTCTAATAAAAAGGATTGC	TGGGAAGTGAGCAAAAGTAAATACA
A35	ACTATCAGTCTCTGCCCTTATTCTG	CACATGTGCATGCTTTTGG
A36	GCCTCACCACCAGAAATCAG	GACATCAGAACTCCCATGAAACT
A37	CCATGATTTTCTTGTGGTGAGA	CACATGTGCTTAGGCCACAAC
A38	GGAGTTCCTGATAACGATTCTGAAG	ACAGAAAAGAAGCCCAGTTGGA
A41	ACATGTGAGGCCATCTCCAC	CCTTACTGTGATGTAGGCACTGTTC

Para a amplificação dos marcadores pertencentes ao Auto 3 foi preparada em primeiro lugar uma mistura de reação, para um volume final de 25µL por amostra, de acordo com o especificado na Tabela 2.5.3.

Tabela 2.5.3 – Reagentes e quantidades respetivas, necessários na preparação da mistura de reação para a amplificação dos marcadores SNP pertencentes ao Auto 3.

Reagente	Quantidade
Tampão PCR (10x)	2,5µL
MgCl ₂ (25mM)	8,0µL
dNTPs (10mM)	7,0µL
Mistura de <i>Primers</i> Auto 3	2,5µL
AmpliTaQ Gold® (5 U/µl)	0,4µL
ADN alvo	4,6µL
Total	25µL

Aquando da preparação da mistura de reação para a amplificação dos marcadores SNP correspondentes ao Auto 3, foi também preparada a mistura de reação para os marcadores pertencentes ao Auto 4, que difere da anterior nos *primers* utilizados, na quantidade de ADN utilizado e no facto de se adicionar água para perfazer o volume final de 25µL por reação, como se pode observar na Tabela 2.5.4

Tabela 2.5.4 – Reagentes e quantidades respetivas, necessários na preparação da mistura de reação para a amplificação dos marcadores SNP pertencentes ao Auto 4.

Reagente	Quantidade
Tampão PCR (10x)	2.5µL
MgCl ₂ (25mM)	8.0µL
dNTPs (10mM)	7.0µL
<i>Primers</i>	2.5µL
AmpliTaQ Gold® (5 U/µl)	0.4µL
Água livre de nucleases	2.3µL
ADN alvo	2.3µL
Total	25µL

Foram realizados dois controlos de amplificação para cada um dos *multiplex* realizados, um controlo positivo com o ADN controlo 9947A (*Applied Biosystems*) e um controlo negativo onde se substituiu o ADN por água.

Ambas as reações de amplificação decorreram num termociclador *GeneAmp PCR system 9700* (*Applied Biosystems*), nas mesmas condições, descritas na Tabela 2.5.5.

Tabela 2.5.5 – Condições de temperatura e tempo para que ocorra cada uma das fases necessárias à amplificação dos marcadores SNP no termociclador.

Temperatura	Tempo	
94°C	5min	
95°C	30s	
60°C	30s	35 ciclos
65°C	30s	
65°C	7min	
4°C	∞	

Após a reação de amplificação, procedeu-se à purificação do produto amplificado, para remover o excesso de *primers* e de dNTPs não utilizados. Para isso retirou-se 2,5µL do produto de PCR para um novo tubo, aos quais se adicionou 1µL de ExoSAP-IT® (*Affymetrix*). Este reagente contém exonucleases que degradam os *primers* e fosfatase alcalina de camarão que degrada os dNTPs. De seguida, as amostras foram incubadas a 37°C durante 45 minutos para a digestão enzimática seguido de 15 minutos a 85°C para inativação das enzimas.

2.5.2. Reação de extensão SBE e purificação do produto sequenciado

A reação de extensão SBE é uma reação de minisequenciação que permite a deteção dos SNP. Esta reação consiste na ligação dos *primers* ao local adjacente do SNP e ocorre de seguida uma extensão pela polimerase de ADN com a ligação da base nucleotídica complementar ao SNP.

Para esta reação é utilizada a metodologia SNaPshot® (*Applied Biosystems*) que utiliza ddNTPs marcados com fluorocromos que se vão ligar à extremidade 3' dos *primers*, isto é, ao local onde ocorre o SNP. Cada ddNTP está assim marcado com diferentes fluorocromos que emitem fluorescência em comprimentos de onda diferentes: guanina emite fluorescência no comprimento de onda azul, a citosina no amarelo, timina no vermelho e adenina no verde.

Para realizar a minisequenciação foram preparadas duas misturas de reação, uma para a análise dos marcadores pertencentes ao Auto 3 e outra para os marcadores pertencentes ao Auto 4. Os *primers* utilizados para esta reação estão descritos na Tabela 2.5.6 e na Tabela 2.5.7. Cada uma das misturas de reação contém SNaPshot®, *primers* e água Milli-Q, estando os volumes necessários descritos na Tabela 2.5.8, para um volume final por reação de 8µL. Foi também preparado um

novo controlo positivo SNaPshot[®], incluído no *kit*, onde se colocou os reagentes necessários nas mesmas quantidades, o ADN controlo fornecido e os *primers* também fornecidos. Os *primers* utilizados foram previamente preparados em alíquotas e armazenados a -20°C.

Tabela 2.5.6 – Primers utilizados para a reação de extensão, respetiva orientação e SNP detetado com esses mesmos primers para o multiplex correspondente ao Auto 3. Adaptado de: Sanchez *et al.*, 2006

SNP ID	SNP	SNP detetado	Orientação de deteção	Local de ligação do <i>primer</i> de extensão
A01	G-A	T/C	Reverse	TCTGAGGCCAGCCAGTT
A06	T-A	A/T	Forward	GTAAGAATTCAAGATGGTATTT
A08	C-T	C/T	Forward	GTTTGTTTATATTATTCTAACTCAA
A10	G-A	T/C	Reverse	CACCGAATTCAACGGGAAG
A12	G-A	A/G	Forward	GTGTTTTTCTAAATCATATTGTCT
A14	A-T	A/T	Forward	AAATACACCCTGAGCTGC
A27	A-G	A/G	Forward	CATCTTGGCAGCATCCTTC
A49	G-A	A/G	Forward	GCAAGAGCCGTGGAATT
A50	C-T	A/G	Reverse	TACAGAAATAACTCTTGCGCTAC
A51	A-G	A/G	Forward	TCCATTCTTTTTTTTTGAAGCCT

Tabela 2.5.7 – Primers utilizados para a reação de extensão, respetiva orientação e SNP detetado com esses mesmos primers para o multiplex correspondente ao Auto 4. Adaptado de: Sanchez *et al.*, 2006

SNP ID	SNP	SNP detetado	Orientação de deteção	Local de ligação do <i>primer</i> de extensão
A04	A-T	T/A	Reverse	TTGTTGACACTTCACCTTCTA
A05	G-A	A/G	Forward	GGCATATCGTATTAAGTGTG
A13	C-T	G/A	Reverse	GATTTTCACAACAACACTTGC
A16	C-A	T/G	Reverse	TCTTGCGGTTACATACATTC
A29	G-A	A/G	Forward	TGTTCTAATAAAAAGGATTGCTCAT
A35	G-A	A/G	Forward	CAGCATACACTCATAGCCAC
A36	T-A	A/T	Forward	CCACCAGAAATCAGGGC
A37	A-G	C/T	Reverse	CTGGCTTTGGCAAGTCC
A38	G-C	G/C	Reverse	CCAGTTGGAGCCTTCTTT
A41	T-A	A/T	Forward	CTCTCTCATGTATCACATTCCT

Tabela 2.5.8 – Reagentes e respectivos volumes necessários à reação de extensão SBE para o Auto 3, Auto 4 e controlo positivo.

Reagente	Quantidade
Produto PCR purificado	1µL
Mistura de reação SNaPshot®	4µL
Primers SBE	1µL
Água Milli-Q	2µL
Total	8µL

De seguida retirou-se 1µL do produto de PCR amplificado e colocou-se num novo tubo e ao qual se adicionou a mistura de reação preparada, correspondente ao Auto 3 e ao Auto 4.

Procedeu-se então à reação de extensão num termociclador *GeneAmp PCR system 9700 (Applied Biosystems)* nas condições apresentadas na Tabela 2.5.9.

Tabela 2.5.9 – Temperaturas e tempos necessários à reação de extensão SBE.

Temperatura	Tempo
96°C	10s
50°C	5s
60°C	30s
4°C	∞

Após esta reação procedeu-se novamente a uma purificação, desta vez apenas com SAP (*Affymetrix*), para remoção dos ddNTPs em excesso. Para isso adicionou-se 1µL de SAP à mistura SBE e incubou-se num termociclador *GeneAmp PCR system 9700 (Applied Biosystems)* durante 85 minutos a 37°C para a enzima atuar, seguido de 15 minutos a 85°C para inativação da enzima.

2.5.3. Eletroforese Capilar do produto SBE

Após a reação de minisequenciação e purificação do produto SBE, procedeu-se à separação desse mesmo produto por eletroforese capilar com deteção por fluorescência.

Para isso procedeu-se em primeiro lugar à preparação das amostras para a eletroforese capilar, numa placa de 96 poços, *MicroAmp[®] Optical* (*Applied Biosystems*). Preparou-se também uma mistura contendo um agente químico desnaturante, neste caso formamida Hi-Di (*Applied Biosystems*) e o marcador de peso molecular interno, o GeneScan[®] 120 LIZ[®] (*Applied Biosystems*). As quantidades necessárias, por amostra, encontram-se descritas na Tabela 2.5.10. A cada poço foi adicionado 18µL da mistura e 2µL do produto SBE, perfazendo um volume total de 20µL.

Tabela 2.5.10 – Reagentes necessários para a preparação do produto SBE para a eletroforese capilar e os respetivos volumes por cada reação.

Reagente	Volume por reação
Formamida Hi-Di	17µL
GeneScan [®] 120 LIZ [®]	1µL
Volume Total	18µL

Seguiu-se uma desnaturação do ADN num termociclador *GeneAmp PCR system 9700* (*Applied Biosystems*) a 95°C durante 3 minutos e a um arrefecimento rápido em placa de gelo novamente por 3 minutos.

Procedeu-se por fim à separação electroforética no sequenciador automático *Genetic Analyzer 3130*, utilizando o polímero POP6[®] (*Applied Biosystems*).

2.5.4. Detecção e análise do produto SBE

A deteção do produto SBE foi realizada com recurso ao *software GeneMapper* versão 3.2 (*Applied Biosystems*), onde se pode observar os picos correspondentes a cada SNP.

2.6. Análise Estatística

A análise estatística dos genótipos obtidos para ambos os tipos de marcadores genéticos estudados foi realizada com recurso a diferentes programas informáticos.

Com recurso ao *software Arlequin*, versão 3.5.1.2 (Excoffier *et al.*, 2005), foram calculadas as frequências alélicas, a heterozigotia observada e esperada, a

probabilidade (p) de desvio do HWE e o *linkage disequilibrium*, para cada *locus*. A probabilidade (p) de desvio do HWE foi calculada segundo o procedimento descrito por Guo e Thompson (1992), tendo sido o valor de p considerado significativo para valores inferiores a 0,05. O cálculo do valor de p para o *linkage disequilibrium* foi realizado segundo o procedimento descrito por Slatkin e Excoffier (1996), e baseia-se na comparação entre a probabilidade calculada quando se considera que não ocorre associação e a probabilidade calculada quando se considera a ocorrência de associação entre *loci*, pelo método de permutação. O valor de p foi considerado significativo para valores inferiores a 0,05. Em ambos os casos foi aplicada a correção de Bonferroni para testes múltiplos, de forma a corrigir erros estatísticos de 1º grau e aumentar o grau de confiança (Bland e Altman, 1995), por divisão do valor de significância pelo número de comparações realizadas. Para o HWE a correção foi realizada para 50 comparações (1 por cada *locus* estudado) e os resultados foram considerados significativos para valores de $p < 0,001$. Para o teste do LD a correção foi realizada para 1 225 comparações realizadas, passando os resultados a serem significativos para valores de $p < 0,00004$.

2.7. Análise da Variação Intrapopulacional e Interpopulacional

Com recurso ao *software Arlequin*, versão 3.5.1.2 (Excoffier *et al.*, 2005) foi também estudada análise da variação molecular (AMOVA do inglês *Analyses of Molecular Variance*) para as populações em estudo, entre as populações em estudo e entre as populações em estudo e a população portuguesa. A análise foi realizada com 16 000 permutações de forma a garantir menos de 1% de diferenças nas probabilidades exatas em 99% dos casos (Excoffier *et al.*, 2005; Guo e Thompson, 1992). O cálculo da variação molecular para os 50 *loci* em estudo foi resumido em tabelas AMOVA que indicam a variabilidade entre as populações (interpopulacional), entre os indivíduos dentro das populações (intrapopulacional) e entre os indivíduos (Excoffier *et al.*, 1992), tendo como base os genótipos observados durante o estudo dos marcadores InDel e dos marcadores SNP. Para a população portuguesa foram utilizados os genótipos obtidos por Matos (2013) e os genótipos disponíveis na base de dados SNPforID. Também nesta análise foi aplicada a correção de Bonferroni, tendo sido o valor de significância corrigido para $p < 0,025$.

2.8. Análise Filogenética

As populações em estudo foram comparadas com outras populações, já caracterizadas para os mesmos marcadores genéticos em estudo, nomeadamente a população de Portugal, Espanha, Alemanha, Dinamarca, China, Somália e Nordeste de Itália. Devido à falta de dados nestas populações, para os 50 marcadores em estudo, foram selecionados dados relativos aos 30 InDel e dados relativos aos 20 SNP, que foram posteriormente compilados numa única matriz de frequências alélicas. Sendo assim foram utilizados relativamente aos 30 InDel: população do Sul de Portugal (Silva *et al.*, 2013), do Centro de Espanha (Martín *et al.*, 2013), da Alemanha (QIAGEN), da Dinamarca (Friis *et al.*, 2012), da República da China (Pepinski *et al.*, 2013), da Somália (Neuvonen *et al.*, 2012) e do Nordeste de Itália (Turrina *et al.*, 2011). Relativamente aos 20 SNP os dados foram retirados da base de dados SNPforID para as populações de Portugal, Espanha, Alemanha, Dinamarca, República da China, Somália e Itália (mais propriamente Veneto no nordeste).

O índice de fixação (F_{st}) é uma medida da diferenciação populacional utilizado aquando o estudo de marcadores polimórficos, como é o caso. Valores de F_{st} pequenos indicam que as frequências alélicas entre as populações são muito semelhantes, enquanto os valores de F_{st} maiores indicam uma maior variação entre as frequências alélicas das populações comparadas. A construção da matriz de distâncias utilizando os valores de F_{st} entre as 8 populações foi realizada com base nas frequências alélicas para cada *locus* estudado e, foi construída com recurso ao *software POPTREE2* (Takezaki *et al.*, 2010). A representação gráfica das relações filogenéticas, tendo como base a matriz construída, foi realizada na plataforma disponível online *Phylogeny.fr*, através do método *neighbor joining* (NJ) implementado no programa *BioNJ* pertencente à plataforma referida (Dereeper *et al.*, 2008; Dereeper *et al.*, 2010; Elias e Lagergren, 2007; Felsenstein, 1989; Gascuel, 1997). A edição da árvore obtida foi posteriormente realizada no *software TreeView*, versão 1.5.2 (Page, 1996).

2.9. Análise dos Parâmetros Forenses

Para os cálculos dos parâmetros forenses, como o poder de discriminação, poder de exclusão, PIC, índice típico de paternidade e probabilidade de *matching* foi

utilizado o *software PowerStats*, versão 12. (Tereba, 1999) Para o cálculo dos parâmetros forenses do conjunto de 50 marcadores estudados, isto é, dos 30 InDel e dos 20 SNP, foi necessário verificar em primeiro lugar a ocorrência de LD, pela posição destes marcadores no genoma, pois estes só podem ser utilizados em investigações forenses se estiverem separados por pelo menos 10Mb. Esta verificação revelou que os marcadores HLD131 e A41 estavam separados por 445 523 nucleótidos apenas, ficando assim o SNP A41 excluído da análise deste painel de marcadores estudado. Sendo assim foram calculadas para os 49 marcadores a probabilidade de *matching* combinada (PMC), o poder de discriminação combinado (PDC), o poder de exclusão combinado (PEC) e o índice de paternidade típico combinado (TPIC), com base nos valores obtidos individualmente no programa *PowerStats*, versão 12.

3. Resultados e Discussão

3.1. Estudo dos Marcadores InDel

Os marcadores InDel tipados com recurso ao *kit* Investigator[®] DIPplex (*QIAGEN*) foram detetados por eletroforese capilar com detecção por fluorescência, sendo o resultado obtido na forma de pico num eletroferograma, representado na Figura 3.1.1. Como se pode observar, a detecção dos 30 InDel e da Amelogenina é realizada apenas em quatro painéis de cores diferentes (azul, verde, preto e vermelho), consoante a marcação fluorescente dos *primers* fornecidos pelo *kit* (o fluorocromo que emite no comprimento de onda amarelo é assinalado como preto). O tamanho dos amplicões gerados na amplificação varia entre 76pb (D83-) e 158pb (D81+) e o tamanho dos picos gerados indicam de forma clara se os alelos estão em heterozigotia (como por exemplo os alelos D93- e D93+) ou em homozigotia (como por exemplo o alelo D99+), salvo algumas exceções.

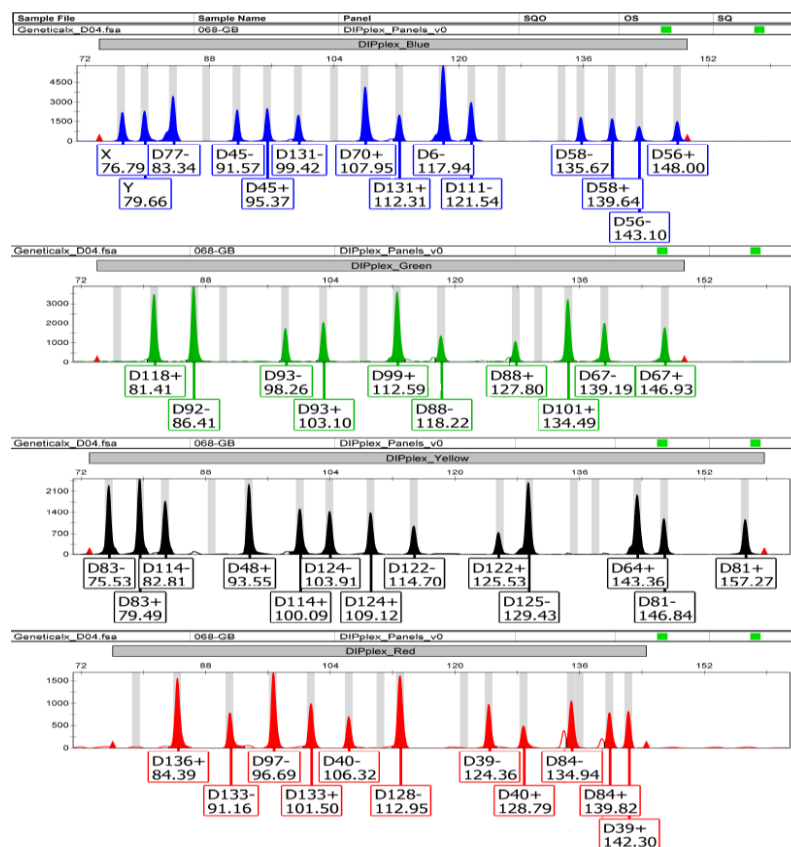


Figura 3.1.1 – Exemplo de um eletroferograma obtido aquando da detecção e análise dos marcadores do tipo InDel: genótipo de um dos indivíduos em estudo (nº68) do sexo masculino oriundo da Guiné-Bissau. Cada painel de cores corresponde à fluorescência emitida pelos fluorocromos, com exceção do preto que corresponde à fluorescência emitida no comprimento de onda amarelo. Cada painel contém 7 a 8 marcadores, o que corresponde a um máximo de 14 a 16 picos.

Do total dos 450 indivíduos estudados com estes marcadores, 23 (5%) apresentaram perfis incompletos com um marcador não amplificado e/ou com grandes desbalanceamentos, nomeadamente nos *loci* HLD70, HLD93, HLD99, HLD88, HLD83, HLD114, HLD124, HLD125, HLD97 e HLD128. A ocorrência de desbalanceamento e *drop-out* nos *loci* HLD83 e HLD97 foi previamente verificada por Fondevila e colaboradores (2011, 2012). Estes autores procederam à sequenciação de forma a entender a causa do fenómeno detetado, porém sem indicação da existência de polimorfismos que possam estar a afetar a ligação dos *primers* e a consequente amplificação do *locus* HLD83. Relativamente ao *locus* HLD97, será discutido em mais pormenor. O desbalanceamento e *drop-out* nos restantes *loci* não se encontram descritos na literatura, porém existem outros casos descritos, como nos *loci* HLD6, HLD40 e HLD58 verificado por Zidkova e colaboradores (2013) na população checa, também no HLD6 e no HLD136 por Kis e colaboradores (2012) na população húngara e, no *locus* HLD131, verificado por Friis e colaboradores (2012) na população dinamarquesa. Apesar destes terem sido verificados apenas em populações europeias, não se deve deixar de lado a hipótese da ocorrência de fenómenos análogos nas populações africanas, principalmente na determinação dos genótipos de indivíduos onde de forma geral a taxa de amplificação foi menor. Deste modo, a ocorrência de *drop-out* pode ficar mascarada pelo tamanho reduzido dos picos detetados, sendo a diferença entre heterozigóticos e homozigóticos muito pequena, pelo que o desbalanceamento não é notório e a ocorrência de *drop-out* não é detetada.

3.1.1. Microvariantes

Durante a análise dos eletroferogramas obtidos foram detetados dois tipos de variações alélicas nos *loci* HLD97, HLD99 e HLD84, não descritas pelo fornecedor do *kit* Investigator[®] DIPplex e que influenciam a interpretação dos genótipos.

3.1.1.1. Locus HLD97

No caso do *locus* HLD97, verificou-se na forma heterozigótica um desbalanceamento do pico correspondente à inserção (D97+) e, na forma homozigótica, uma diminuição do tamanho do pico para esse mesmo alelo e até mesmo ausência de pico, comparativamente aos restantes alelos homozigóticos. Na Figura 3.1.2 pode-se observar dois exemplos da microvariante para o alelo D97+ (a e

b) e o exemplo de um indivíduo onde a detecção do alelo ocorreu sem qualquer alteração e de forma balanceada (c).

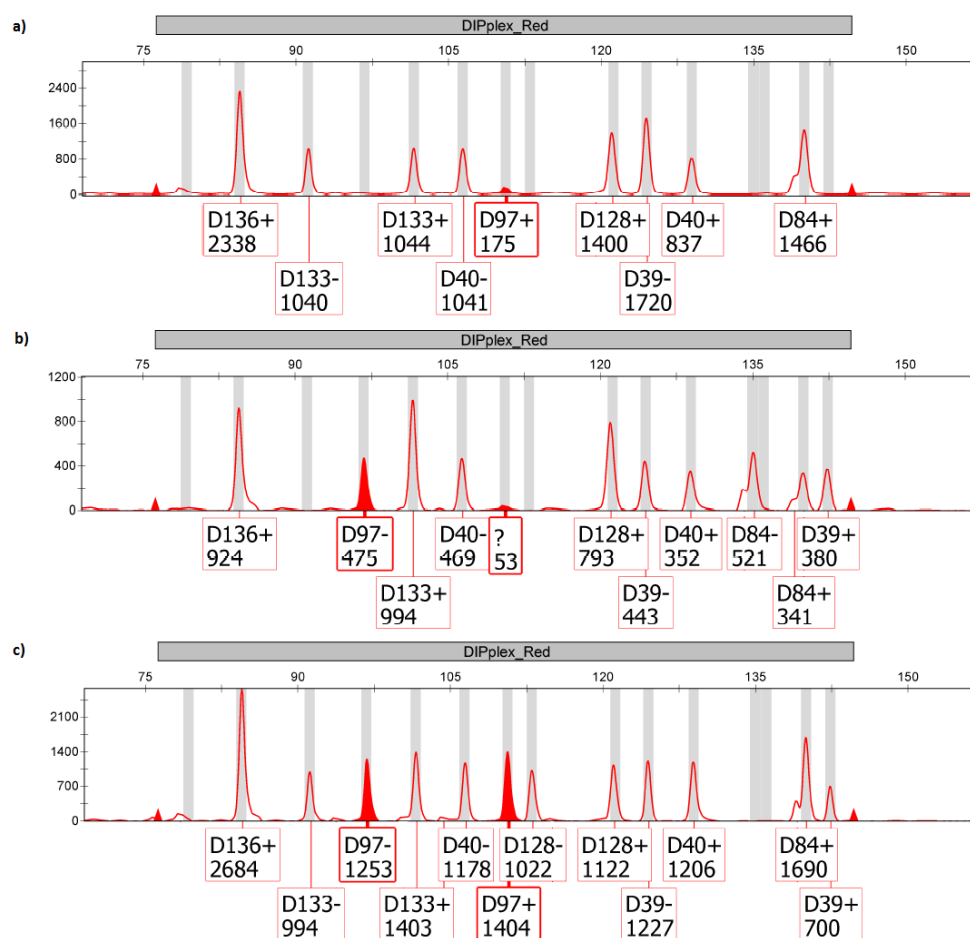


Figura 3.1.2 – Diferenças no tamanho (RFUs) dos picos do *locus* correspondente à inserção do marcador HLD97 (D97+ marcado a vermelho): a) pode-se verificar a presença de um pico com apenas 175 RFUs no local correspondente ao alelo e a ausência de pico para o alelo correspondente à deleção (D97-), indicando que este indivíduo é homocigótico para a inserção neste *locus*, apesar do pico ter um tamanho muito menor que os restantes detetados no painel; b) pode-se verificar a presença de um pico para o alelo correspondente à deleção (D97-) com um tamanho de 475 RFUs, similar aos picos para os alelos heterocigóticos. Ao analisar com maior atenção, verificou-se a presença de um pico (não detetado pelas definições estabelecidas) com apenas 53 RFUs para o alelo correspondente à inserção, sugerindo que o indivíduo seja heterocigótico e não homocigótico para o *locus* em questão; c) exemplo de um indivíduo heterocigótico para o marcador em questão e cuja detecção dos alelos correspondentes à deleção e à inserção ocorreu sem problemas e com um tamanho balanceado.

Dos 83 indivíduos de origem guineense estudados, só não foi possível confirmar o genótipo para este *locus* num indivíduo, com total ausência de pico para ambos os alelos. Já dos 367 indivíduos de origem cabo-verdiana estudados, em 3 não foi possível verificar o genótipo para o *locus* em questão e, em 13 não foi possível

determinar a presença do alelo D97+. No total dos 450 indivíduos, verificou-se uma frequência de 0,023 desta microvariante. A verificação dos genótipos deste *locus* nos indivíduos referidos foi realizada por repetição da genotipagem, onde se verificou que esta microvariante era reprodutível, sem esclarecer os genótipos reais. Para além disso, a genotipagem dos filhos destes indivíduos também revelou que este marcador era transmitido à descendência.

Uma das formas de confirmar o genótipo dos indivíduos e estudar a microvariante é por sequenciação da região do alelo em questão. Fondevila e colaboradores (2011), Friis e colaboradores (2012) e Martín e colaboradores (2013) procederam à sequenciação dessa região, de forma a entender a causa da redução do tamanho do pico para o alelo D97+, por suspeita da presença de um polimorfismo que afecta o local de ligação dos *primers*, impedindo a correta amplificação desse alelo. Estes autores verificaram a presença de um SNP A-G (rs17245568) localizado 61pb a jusante do *locus*, isto é, entre os 97pb e os 110pb do frangmento de PCR e que os fragmentos que possuíam o alelo A apresentavam uma taxa de amplificação menor que os fragmentos que possuíam o alelo G, estando assim a presença da adenina associada à deteção de um pico menor e desbalanceado correspondente ao alelo D97+.

Em dois estudos de avaliação do *kit* Investigator®DIPplex com quatro populações dos Estados Unidos (afro-americana, caucasiana, hispânica e asiática), Fondevila e colaboradores (2011, 2012) observaram uma maior frequência desta microvariante na população afro-americana (0,22) e menor na população caucasiana (0,044), apenas na forma heterozigótica. Na forma homozigótica, estes autores estimam a presença da microvariante em cerca de 5% da população de origem africana. Também LaRue e colaboradores (2012), num estudo de validação do *kit*, com recurso às mesmas populações dos Estados Unidos, verificaram uma maior frequência da microvariante na população afro-americana (0,280) mas não observaram esta microvariante na população caucasiana. Neste estudo, a frequência da microvariante presente em 8% dos 450 indivíduos tipados, revelou-se muito menor, porém só foram contabilizados os alelos onde não foi possível determinar o genótipo, desconsiderando os desbalanceamentos aquando da deteção dos dois picos para o *locus*.

3.1.1.2. Loci HLD99 e HLD84

Relativamente às microvariantes móveis, foram detetados dois picos *off ladder* que apresentavam a mesma força de sinal que os picos para os alelos constituintes dos loci HLD99 e HLD84 (ver Figura 3.1.3).

No caso do locus HLD99, verificou-se que o pico correspondente ao alelo para a inserção apresentava um nucleótido a menos do que o descrito, isto é cerca de 112pb em vez de 113pb, tendo sido denominado de D99+(-1). No caso do locus HLD84, verificou-se que o pico correspondente ao alelo para inserção apresentava quatro nucleótido a menos do que o descrito, isto é, cerca de 136pb em vez de 140pb, tendo sido denominado de D84+(-4).

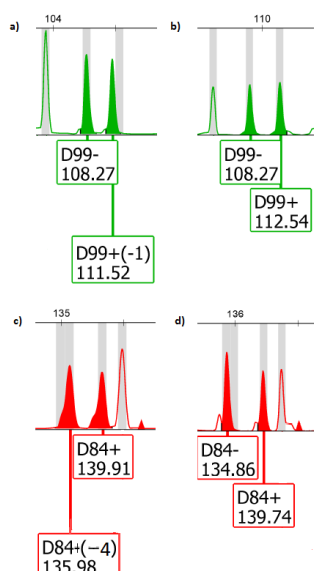


Figura 3.1.3 – Microvariantes detetadas aquando da deteção e análise dos marcadores InDel: a) pico *off ladder* assinalado como D99+(-1), detetado no painel correspondente à cor verde, com um nucleótido a menos do que b) o alelo D99+ descrito; c) pico *off ladder*, assinalado como D84+(-4), detetado no painel correspondente à cor vermelha, com quatro nucleótido a menos do que d) o alelo D84+ descrito.

A verificação destes alelos não descritos foi realizada por repetição da tipagem e pelo estudo dos filhos dos indivíduos onde foram detetadas as microvariantes, o que revelou novamente a reprodutibilidade e a transmissão à descendência destes alelos. Também neste caso, a sequenciação das regiões onde se encontram estes alelos seria a melhor forma de estudar e caracterizar ambas as microvariantes.

Nos mesmos estudos de avaliação e validação do *kit*, Fondevila e colaboradores (2011, 2012) identificaram e caracterizaram estas microvariantes. No caso da microvariante no *locus* HLD99, a sequenciação da região revelou a existência de deleção de uma adenina, 4pb a montante do alelo para a inserção, identificada com o número rs11346981. No caso do *locus* HLD84, revelou uma deleção de 4 bases (ATTA), localizada 10pb a jusante do alelo para a inserção. Essa deleção está identificada com o número rs11573892. Ambas as microvariantes são também polimorfismos InDel localizados nas regiões em estudo, que dão origem a um terceiro alelo de tamanho diferente.

A presença de um terceiro alelo para estes *loci* permite uma maior discriminação entre indivíduos, isto é, permite o aumento do poder de discriminação do *kit* Investigator[®]DIPplex (*QIAGEN*) (Fondevila *et al.*, 2011b; Fondevila *et al.*, 2012; LaRue *et al.*, 2012). Por esta razão, estes alelos foram considerados como pertencentes aos marcadores em estudo e incluídos no estudo e caracterização das populações imigrantes de origem guineense e cabo-verdiana a residir na região de Lisboa.

3.2. Estudo dos Marcadores SNP

A deteção e análise, dos 20 marcadores SNP, foi realizada por eletroforese capilar com deteção por fluorescência. Na Figura 3.2.1 é apresentado um exemplo de um eletroferograma obtido, para ambos os *multiplex* realizados, sendo o primeiro painel relativo ao Auto 3 e o segundo ao Auto 4. Cada pico está identificado com o SNP ao qual corresponde e com o nucleótido que o constitui.

Ao contrário dos InDel, estes marcadores são mais simples de analisar por se encontrarem agrupados ao longo de cada painel, mas por outro lado a obtenção de perfis completos é mais difícil e está mais sujeita a erros e contaminações. Ainda comparativamente aos marcadores InDel, verifica-se a presença de um maior desbalanceamento dos picos correspondentes aos alelos heterozigóticos e um maior número de alelos não amplificados ou com sinal de deteção muito baixo do que se verifica aquando a análise dos marcadores InDel. Foi nos *loci* pertencentes ao Auto 4 onde se verificou um maior desbalanceamento, nomeadamente no A41, A13, A29, A16, A05 e A04. Os *loci* detetados através da reação de PCR *multiplex* Auto 3

revelaram-se mais difíceis de detetar, com ausência de pico para diversos marcadores nos indivíduos estudados.

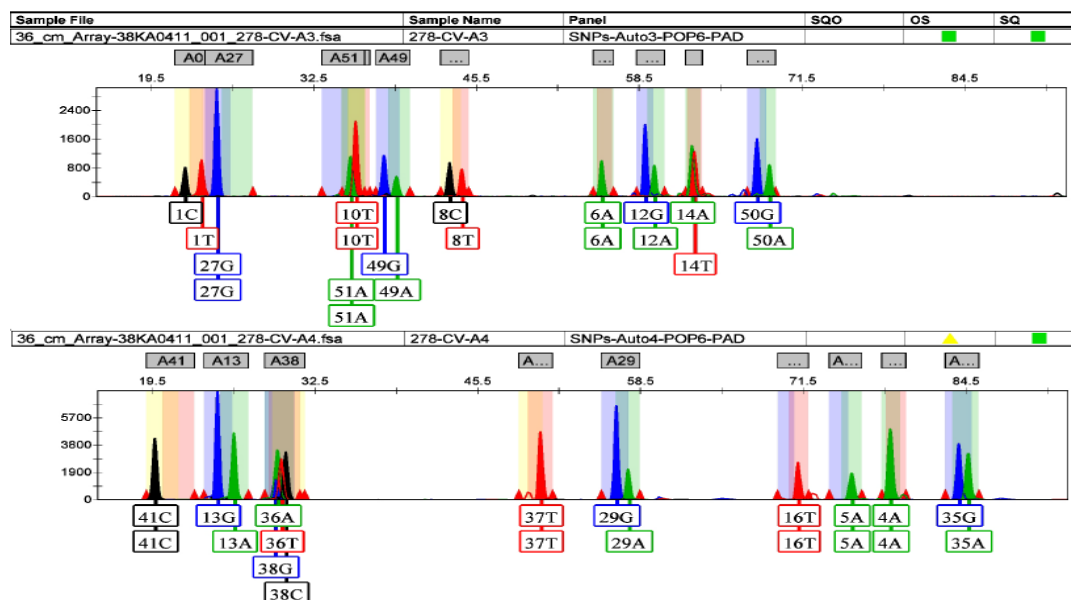


Figura 3.2.1 – Exemplo de um eletroferograma obtido aquando da deteção e análise dos marcadores do tipo SNP: genótipo de um dos indivíduos em estudo (nº278) oriundo de Cabo Verde. O primeiro painel corresponde ao *multiplex* Auto 3 (A3) e o segundo ao *multiplex* Auto 4 (A4), cada um com 10 marcadores, representados no topo de cada painel.

Børsting e colaboradores (2009) num estudo de validação do SNPforID 52-plex, optaram por excluir três marcadores devido ao desbalanceamento e à difícil deteção, entre os quais o A13 com um sinal do produto SBE fraco e o A35 devido ao tamanho do pico para o alelo G ser muito superior ao pico para o alelo A. Apesar de não excluídos do *multiplex*, estes autores verificaram ainda um silenciamento parcial nos *loci* A16 e A41 sem mutações associadas, pois testaram a amplificação do *locus* A16 com novos *primers* que revelou heterozigotia. Neste estudo o desbalanceamento do *locus* A35 não foi tão notório porém foi verificado o fenómeno noutros *loci* que não foram visualizados por outros autores pelo que podem ser característicos das populações em estudo ou devidos a uma taxa de amplificação menor do que a esperada, sem amplificação desses alelos. Também Dario e colaboradores (2009) não verificaram qualquer mutação associada a estes 20 *loci* SNP em estudo.

A tipagem dos 450 indivíduos em estudo revelou maior dificuldade na obtenção de perfis completos, tendo sido essa maior no painel correspondente aos

marcadores pertencentes ao Auto 3. Destes 450 indivíduos, apenas 79% apresentaram perfis completos. Nos restantes 21% não houve detecção de sinal para pelo menos um marcador ou observou-se um elevado desbalanceamento que impede a correta identificação do genótipo. Diferentes autores relevaram também a obtenção de perfis incompletos (após a tipagem em duplicado para cada indivíduo), como Børsting e colaboradores (2009) onde apenas 78% dos indivíduos estudados apresentaram perfis completos; Sanchez e colaboradores (2006) em 20% dos 700 indivíduos estudados; Montelius e colaboradores (2009) em 30% dos 162 indivíduos e Santos e colaboradores (2011), que em 46 indivíduos, obtiveram perfis completos em 99,87% da população.

Aquando da análise dos eletroferogramas obtidos verificou-se também um sinal mais fraco dos picos correspondentes aos alelos constituídos por uma adenina quando estes se encontravam em heterozigotia, não tendo sido esses alelos assinalados, pois na metodologia SNaPshot[®] (*Applied Biosystems*) utilizada ocorre uma tendência para adicionar adeninas ao produto não específico, gerado durante a amplificação, que provoca interferências na detecção do sinal emitido no comprimento de onda verde (Børsting *et al.*, 2009; Fondevila *et al.*, 2013; Sanchez *et al.*, 2008)

Contrariamente ao verificado para os *loci* InDel, os problemas associados à detecção e análise destes *loci* SNP não estão associados a nenhuma população em específico nem a mutações ou polimorfismos presentes no local de ligação dos *primers* utilizados.

3.3. Caracterização Genética das Populações em Estudo

O estudo das frequências alélicas de uma população é uma ferramenta muito útil na área forense, pois permite a construção de bases de dados populacionais que irão servir depois para inferir acerca das frequências dos perfis de ADN obtidos durante uma investigação (Butler, 2010b). Para além disso, o conhecimento das frequências alélicas características de uma população pode permitir avaliar a origem do indivíduo a quem pertence a amostra, caso se verifique a presença de AIMs (Kosoy *et al.*, 2009; Rosenberg *et al.*, 2005), apesar de se verificar muitas vezes maiores variações genéticas entre indivíduos do que entre regiões (Rosenberg *et al.*, 2002).

A caracterização de uma população é assim realizada pela contagem da ocorrência dos alelos, pelo que quanto maior o número de indivíduos em estudo, mais fidedigna será a base de dados construída. O uso destas, juntamente com a análise do HWE irá permitir ao especialista forense o cálculo da frequência de um perfil genético, pois quando uma população se encontra em HWE, as frequências genotípicas das gerações seguintes podem ser previstas pelas frequências alélicas obtidas anteriormente. Quando uma população apresenta frequências genotípicas observadas semelhantes às esperadas, calculadas pelas frequências alélicas, a população encontra-se em HWE, requerimento necessário para ser considerada uma população ideal e representativa (Goodwin *et al.*, 2007c).

Sendo o número de imigrantes de origem guineense e cabo-verdiana em Portugal elevado, principalmente na região de Lisboa, é necessária a construção de uma base de dados para estas populações, cada vez mais frequentes no SGBF do INMLCF, com recurso a outros marcadores genéticos além dos STRs correntemente utilizados.

Sendo assim, para a caracterização das populações em estudo foram calculados diversos parâmetros, entre os quais: as frequências alélicas, a frequência de heterozigóticos e homozigóticos, a heterozigotia observada e esperada e o valor de significância (p) para o HWE.

3.3.1. Estudo da População Guineense residente na região de Lisboa

Os resultados obtidos dos diversos parâmetros calculados para a população guineense imigrante a residir na região de Lisboa encontram-se descritos na Tabela 3.3.1.

Foram tipados 83 indivíduos representativos da população em estudo, porém os marcadores InDel HLD83, HLD125 e HLD97 e os marcadores SNP são apresentados em menor número, por apresentarem um elevado desbalanceamento entre os picos heterozigóticos e/ou por não terem sido detetados.

As microvariantes detetadas nos *loci* HLD99 e HLD84 foram consideradas como um terceiro alelo por apresentarem uma frequência superior a 1% na população e por permitirem um aumento do poder de discriminação do painel de marcadores estudados.

Tabela 3.3.1 – Dados relativos aos 50 loci estudados da população guineense residente na região de Lisboa: número de alelos analisados (2N), frequências alélicas, frequências de heterozigotia (H) e homozigotia (h), heterozigotia observada (Ho) e esperada (He) e o valor de *p* para o HWE.

<i>Locus</i>	2N	Frequências Alélicas			H	h	Ho	He	<i>p-value</i>
		Deleção	Inserção	Microvariante					
HLD77	166	0,71084	0,28916	-	38,60%	61,40%	0,38554	0,41358	0,59483
HLD45	166	0,59036	0,40964	-	43,40%	56,60%	0,43373	0,48660	0,36582
HLD131	166	0,17470	0,82530	-	25,30%	74,70%	0,25301	0,29011	0,25710
HLD70	166	0,09639	0,90361	-	19,30%	80,70%	0,19277	0,17525	1,00000
HLD6	166	0,56627	0,43373	-	53,00%	47,00%	0,53012	0,49419	0,65564
HLD111	166	0,66867	0,33133	-	42,20%	57,80%	0,42169	0,44578	0,62823
HLD58	166	0,89759	0,10241	-	20,50%	79,50%	0,20482	0,18496	0,59172
HLD56	166	0,45181	0,54819	-	47,00%	53,00%	0,46988	0,49836	0,65893
HLD118	166	0,76506	0,23494	-	39,80%	60,20%	0,39759	0,36166	0,54041
HLD92	166	0,68072	0,31928	-	37,30%	62,70%	0,37349	0,43731	0,20820
HLD93	166	0,51205	0,48795	-	56,60%	43,40%	0,56627	0,50274	0,27771
HLD99	166	0,26506	0,69880	0,03615	42,20%	57,80%	0,42169	0,44279	0,49124
HLD88	166	0,32530	0,67470	-	41,00%	59,00%	0,40964	0,44162	0,61517
HLD101	166	0,15663	0,84337	-	24,10%	75,90%	0,24096	0,26579	0,40712
HLD67	166	0,30723	0,69277	-	37,30%	62,70%	0,37349	0,42826	0,30196
HLD83	165	0,29697	0,70303	-	35,40%	64,60%	0,35366	0,42159	0,18432
HLD114	166	0,23494	0,76506	-	37,30%	62,70%	0,37349	0,36166	1,00000
HLD48	166	0,16265	0,83735	-	30,10%	69,90%	0,30120	0,27404	0,68245
HLD124	166	0,84940	0,15060	-	27,70%	72,30%	0,27711	0,25739	0,68202
HLD122	166	0,60241	0,39759	-	50,60%	49,40%	0,50602	0,48193	0,81753
HLD125	164	0,76829	0,23171	-	41,50%	58,50%	0,41463	0,35822	0,21467
HLD64	166	0,21687	0,78313	-	28,90%	71,10%	0,28916	0,34173	0,19455
HLD81	166	0,40964	0,59036	-	50,60%	49,40%	0,50602	0,48660	0,82033
HLD136	166	0,16867	0,83133	-	26,50%	73,50%	0,26506	0,28215	0,69082
HLD133	166	0,58434	0,41566	-	51,80%	48,20%	0,51807	0,48872	0,65085

Tabela 3.3.1 – Continuação – Dados relativos aos 50 *loci* estudados da população guineense residente na região de Lisboa: número de alelos analisados (2N), frequências alélicas, frequências de heterozigotia (H) e homozigotia (h), heterozigotia observada (Ho) e esperada (He) e o valor de *p* para o HWE.

<i>Locus</i>	2N	Frequências Alélicas			H	h	Ho	He	<i>p-value</i>
		Deleção	Inserção	Microvariante					
HLD97	164	0,59756	0,40244	-	22,00%	78,00%	0,21951	0,48391	0,00000
HLD40	166	0,72892	0,27108	-	42,20%	57,80%	0,42169	0,39759	0,78104
HLD128	166	0,27108	0,72892	-	37,30%	62,70%	0,37349	0,39759	0,58638
HLD39	166	0,45783	0,54217	-	55,40%	44,60%	0,55422	0,49945	0,37656
HLD84	166	0,26506	0,68072	0,05422	41,00%	59,00%	0,40964	0,46623	0,18467
		Alelo 1	Alelo 2	-					
A27	144	(G) 0,32192	(A) 0,67808	-	45,80%	54,20%	0,45833	0,44279	1,00000
A01	158	(C) 0,54375	(T) 0,45625	-	45,60%	54,40%	0,45570	0,49923	0,49816
A06	140	(A) 0,56738	(T) 0,43262	-	44,30%	55,70%	0,44286	0,49527	0,46540
A08	140	(C) 0,57042	(T) 0,42958	-	54,30%	45,70%	0,54286	0,49332	0,46622
A10	132	(C) 0,66165	(T) 0,33835	-	33,30%	66,70%	0,33333	0,44784	0,05123
A12	144	(G) 0,31973	(A) 0,68027	-	43,10%	56,90%	0,43056	0,43269	1,00000
A14	160	(A) 0,69565	(T) 0,30435	-	46,30%	53,80%	0,46250	0,42759	0,60044
A49	136	(G) 0,54286	(A) 0,45714	-	47,10%	52,90%	0,47059	0,50196	0,63471
A50	136	(G) 0,52174	(A) 0,47826	-	52,90%	47,10%	0,52941	0,50327	0,80791
A51	136	(G) 0,44928	(A) 0,55072	-	52,90%	47,10%	0,52941	0,49673	0,62991
A04	156	(A) 0,60377	(T) 0,39623	-	50,00%	50,00%	0,50000	0,47932	0,81326
A05	158	(G) 0,69565	(A) 0,30435	-	50,60%	49,40%	0,51250	0,42759	0,11337
A37	160	(C) 0,51553	(T) 0,48447	-	70,00%	30,00%	0,70000	0,50283	0,00064
A35	158	(A) 0,73750	(G) 0,26250	-	40,50%	59,50%	0,40506	0,39281	1,00000
A29	154	(G) 0,67089	(A) 0,32911	-	66,20%	33,80%	0,66234	0,44589	0,00000
A36	156	(A) 0,25786	(T) 0,74214	-	46,20%	53,80%	0,46154	0,38379	0,08025
A41	162	(T) 0,56173	(C) 0,43827	-	58,00%	42,00%	0,58025	0,49544	0,17497
A38	156	(G) 0,63522	(C) 0,36478	-	53,80%	46,20%	0,53846	0,47014	0,23248
A16	146	(G) 0,84416	(T) 0,15584	-	27,40%	72,60%	0,27397	0,27662	1,00000
A13	160	(G) 0,81988	(A) 0,18012	-	36,30%	63,80%	0,36250	0,29866	0,06201

Relativamente às frequências alélicas calculadas para os 30 marcadores InDel em estudo, pode-se observar no Gráfico 3.3.1 um elevado desbalanceamento entre as frequências dos alelos correspondentes à deleção e à inserção, com destaque para os *loci* HLD70 e HLD58. No caso do *locus* HLD70, onde a deteção do sinal ocorreu também de forma desbalanceada, o alelo para a deleção apresenta uma frequência de 0,096 (mínima) e o alelo para a inserção de 0,904 (máxima). Esta diferença pode dever-se a um *drop-out* do alelo para a deleção (D70-), não descrito na literatura. No caso do *locus* HLD58 o alelo para a deleção apresenta uma frequência de 0,898 e o alelo para a inserção de 0,102, contudo a deteção dos alelos decorreu na normalidade nesta população. Zidkova e colaboradores (2013) identificaram a ocorrência de *drop-out* no *locus* HLD58, não podendo esta causa ser desconsiderada para a baixa frequência do alelo para a inserção (D58+). Apenas os *loci* HLD56, HLD93 e HLD39 apresentam frequências alélicas próximas dos 0,500.

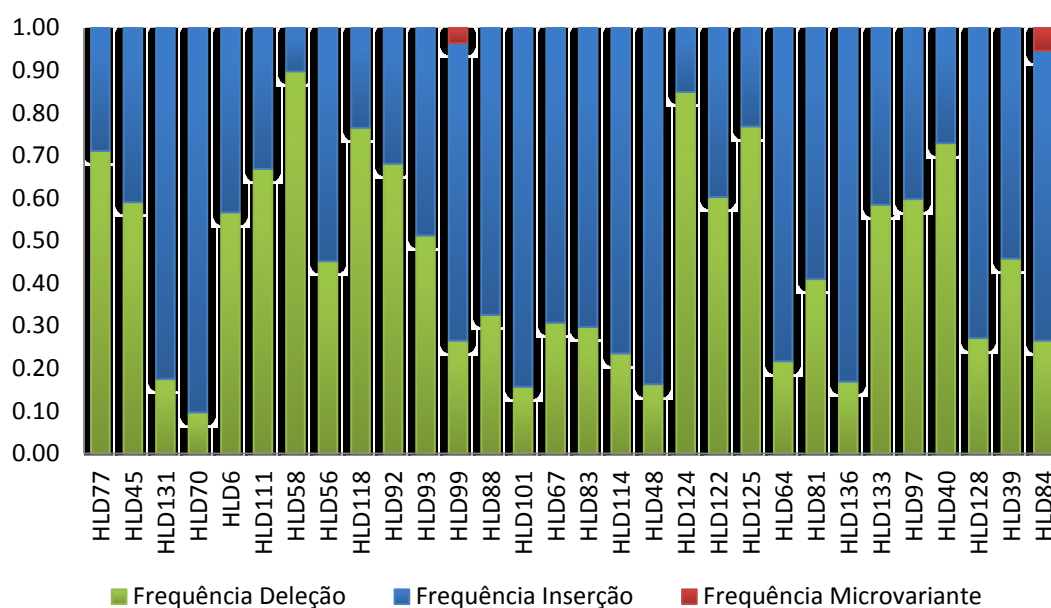


Gráfico 3.3.1 – Frequências alélicas para os 30 marcadores InDel em estudo na população guineense imigrante na região de Lisboa.

Quanto às microvariantes detetadas, o alelo D99+(-1) foi detetado 6 vezes, em heterozigotia, tendo uma frequência de 0,036 e o alelo D84+(-4) foi detetado 9 vezes em 8 indivíduos (1 indivíduo apresentava homozigotia para o alelo em questão), apresentado uma frequência de 0,054.

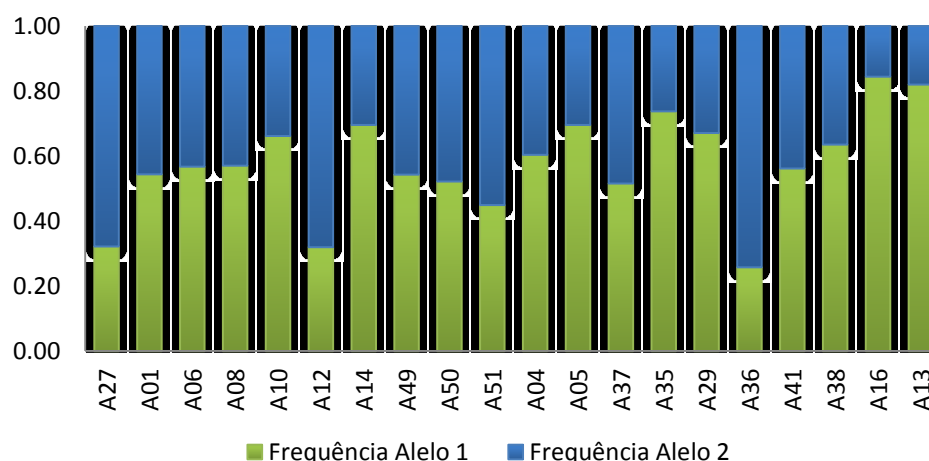


Gráfico 3.3.1 – Frequências alélicas para os 20 marcadores SNP em estudo na população guineense migrante na região de Lisboa. A identificação dos alelos encontra-se na Tabela 3.3.1.

As frequências alélicas para os 20 marcadores SNP estudados encontram-se representadas graficamente no Gráfico 3.3.2, onde se pode observar novamente um desbalanceamento, com frequências alélicas entre 0,1 e os 0,8, nomeadamente nos *loci* A16 e A13. No *locus* A16 o alelo 1 constituído por uma guanina tem uma frequência de 0,844 (máxima) e o alelo 2 constituído por uma timina tem uma frequência de 0,156 (mínima). O pico detetado para o alelo 2 (T) apresentou-se, de forma geral, menor que o pico para o alelo 1 (G), podendo ter ocorrido *drop-out* (Børsting *et al.*, 2009) do alelo 2 e este não ser detetado em alguns indivíduos. No *locus* A13, o alelo 1 também constituído por uma guanina tem uma frequência de 0,820 e o alelo 2 constituído por uma adenina tem uma frequência na população de 0,180. Novamente a deteção destes alelos decorreu de forma desbalanceada, com diminuição do tamanho do pico correspondente ao alelo 2 (A), podendo, tal como referido para o *locus* A16, a ocorrência de *drop-out* deste alelo influenciar a determinação do genótipo e como consequência a frequência do alelo nesta população. De notar que o mesmo autor que reportou o desbalanceamento no *locus* A16, identificou também um sinal de deteção fraco associado a este *locus*. Apesar disso, pode-se sugerir que estas frequências sejam também características da população guineense, porém seria necessário aumentar o número de indivíduos em

estudo. Apenas em 8 dos 20 *loci* SNP verificaram-se frequências alélicas próximas dos 0,5, como o caso do A01, A06, A08, A49, A50, A51, A37 e A41.

Estes 50 *loci* estudados, apresentam uma frequência de homozigotia média de 57,79% e uma frequência de heterozigotia média de 42,21%, o que significa que existem mais alelos na forma homozigótica do que heterozigótica e que a ocorrência na forma homozigótica é superior. Apesar destes dados, verificou-se a inexistência de genótipos homozigóticos para a deleção no *locus* HLD70, para a inserção no *locus* HLD58 tal como nos alelos constituídos por uma adenina nos *loci* A29 e A13. Estes alelos apresentam uma baixa frequência na população em estudo e foram detetados com desbalanceamento e com possibilidade de *drop-out* sendo essa a razão de apenas terem sido observados em heterozigotia. De notar que, com exceção do *locus* A29, os genótipos para estes *loci* são na sua maioria homozigóticos (80,7% no HLD70, 79,5% no HLD58 e 63,8% no A13).

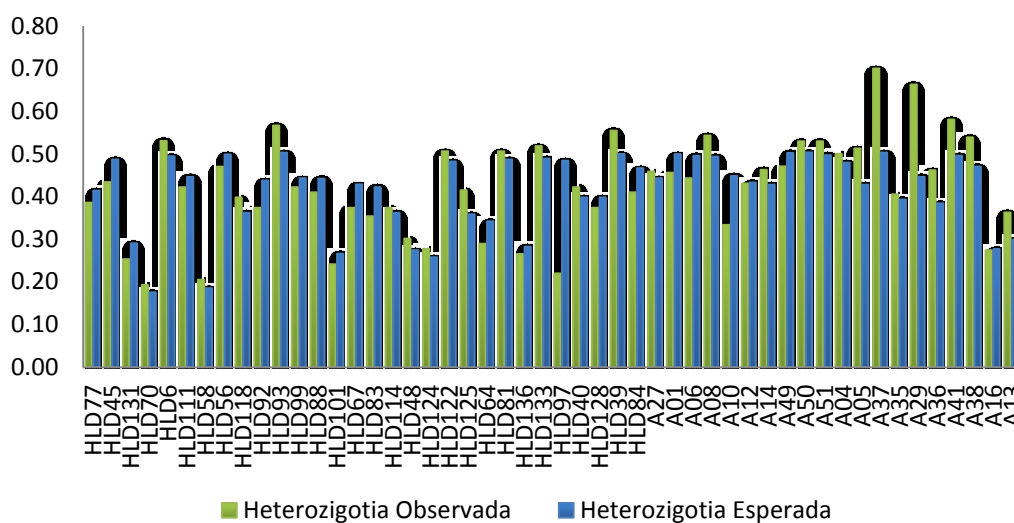


Gráfico 3.3.2 – Heterozigotias observadas e esperadas para os 50 marcadores InDel e SNP estudados na população guineense imigrante residente na região de Lisboa.

As heterozigotias observadas e esperadas para todos os 50 marcadores podem ser observadas na Gráfico 3.3.2. A maioria dos *loci* estudados apresentam uma diferença muito pequena entre a H_o e a H_e e, de uma forma muito geral, os valores de H_o encontram-se entre 0,19 e 0,70, com uma média de 0,42 e os valores de H_e encontram-se entre 0,17 e 0,50, com uma média de 0,41. Os *loci* HLD97, e A10

apresentam uma H_o consideravelmente menor do que a H_e , enquanto os *loci* A05, A37, A29, A41, A38 e A13 apresentam uma H_o consideravelmente maior do que a H_e . Ainda relativamente ao *locus* HLD97, este apresenta um valor de H_o de 0,220, que é cerca de metade da H_e para este *locus* (0,484). Estas diferenças refletem-se nos valores de p para o HWE, apresentando os *loci* HLD97 e A29 um valor de $p=0,00000$ e o *locus* A37 um $p=0,00064$. Após aplicar a correção de Bonferroni ($p<0,001$) o valor de p para estes marcadores manteve-se significativo, indicando um desvio do HWE. Todos os restantes *loci* encontram-se em HWE ($0,062<p<1$).

A presença de alelos nulos para os *loci* HLD97 (2N=164) e A29 (2N=154) e o reduzido número da população em estudo pode estar na base no desvio do HWE para estes *loci*. Ainda no caso do *locus* HLD97, a diferença entre a frequência de alelos homozigóticos (78%) e heterozigóticos (22%) e o facto de a H_o ser metade da H_e , ambas devidas à microvariante detetada, pode explicar o desvio do HWE. Em ambos os casos, o aumento do número de indivíduos estudados e uma otimização das reações de amplificação, poderá resolver este desvio.

3.3.2. Estudo da População Cabo-verdiana residente na região de Lisboa

Os resultados obtidos dos diversos parâmetros calculados para a população cabo-verdiana imigrante a residir na região de Lisboa encontram-se descritos na Tabela 3.3.2.

Foram tipados um total de 367 indivíduos representativos da população em questão. Tal como ocorreu com a população guineense, alguns marcadores não foram amplificados, nomeadamente os marcadores SNP, e os InDel HLD70, HLD93, HLD99, HLD88, HLD114, HLD124, HLD125, HLD97 e HLD128, o que se reflete num número total, de alelos estudados, menor para esses marcadores (ver Tabela 3.3.2).

Tabela 3.3.2 – Dados relativos aos 50 loci estudados da população cabo-verdiana residente na região de Lisboa: número de alelos analisados (2N), frequências alélicas, frequências de heterozigotia (H) e homozigotia (h), heterozigotia observada (Ho) e esperada (He) e o valor de *p* para o HWE.

<i>Locus</i>	2N	Frequências Alélicas			H	h	Ho	He	<i>p-value</i>
		Deleção	Inserção	Microvariante					
HLD77	734	0,66894	0,33106	-	44,40%	55,60%	0,44414	0,44352	1,00000
HLD45	734	0,57221	0,42779	-	46,30%	53,70%	0,46322	0,49024	0,33428
HLD131	734	0,22207	0,77793	-	30,20%	69,80%	0,30245	0,34598	0,02326
HLD70	733	0,15825	0,84175	-	25,70%	74,30%	0,25683	0,26708	0,44023
HLD6	734	0,55995	0,44005	-	49,30%	50,70%	0,49319	0,49349	1,00000
HLD111	734	0,60218	0,39782	-	53,40%	46,60%	0,53406	0,47977	0,03779
HLD58	734	0,82289	0,17711	-	30,00%	70,00%	0,29973	0,29188	0,72190
HLD56	734	0,35831	0,64169	-	47,70%	52,30%	0,47684	0,46048	0,57348
HLD118	734	0,67439	0,32561	-	45,50%	54,50%	0,45504	0,43978	0,55029
HLD92	734	0,67847	0,32153	-	45,20%	54,80%	0,45232	0,43689	0,54772
HLD93	733	0,57435	0,42565	-	53,00%	47,00%	0,53005	0,48978	0,13588
HLD99	732	0,28415	0,70628	0,00956	46,20%	53,80%	0,46175	0,42090	0,15523
HLD88	732	0,36749	0,63251	-	46,60%	53,40%	0,46575	0,46532	1,00000
HLD101	734	0,27384	0,72616	-	41,10%	58,90%	0,41144	0,39825	0,59992
HLD67	734	0,36104	0,63896	-	49,90%	50,10%	0,49864	0,46201	0,14386
HLD83	733	0,38881	0,61119	-	38,30%	61,70%	0,38251	0,47555	0,00017
HLD114	733	0,31105	0,68895	-	39,10%	60,90%	0,39071	0,42847	0,11031
HLD48	734	0,29973	0,70027	-	43,10%	56,90%	0,43052	0,42035	0,70725
HLD124	732	0,64891	0,35109	-	47,30%	52,70%	0,47268	0,45628	0,56576
HLD122	734	0,61444	0,38556	-	46,60%	53,40%	0,46594	0,47445	0,74099
HLD125	733	0,65075	0,34925	-	45,90%	54,10%	0,45902	0,45546	0,90953
HLD64	734	0,26703	0,73297	-	37,60%	62,40%	0,37602	0,39198	0,42507
HLD81	734	0,48501	0,51499	-	48,50%	51,50%	0,48501	0,50023	0,59780
HLD136	734	0,22752	0,77248	-	33,50%	66,50%	0,33515	0,35199	0,37476
HLD133	734	0,53678	0,46322	-	47,40%	52,60%	0,47411	0,49797	0,39899

Tabela 3.3.2 – Continuação - Dados relativos aos 50 loci estudados da população cabo-verdiana residente na região de Lisboa: número de alelos analisados (2N), frequências alélicas, frequências de heterozigotia (H) e homozigotia (h), heterozigotia observada (Ho) e esperada (He) e o valor de *p* para o HWE.

<i>Locus</i>	2N	Frequências Alélicas			H	h	Ho	He	<i>p-value</i>
		Deleção	Inserção	Microvariante					
HLD97	716	0,56783	0,43217	-	40,70%	59,30%	0,40741	0,49354	0,00096
HLD40	734	0,67984	0,32016	-	40,10%	59,90%	0,40054	0,43591	0,12228
HLD128	732	0,37705	0,62295	-	43,70%	56,30%	0,43716	0,47041	0,18385
HLD39	734	0,51635	0,48365	-	52,00%	48,00%	0,52044	0,50015	0,46454
HLD84	734	0,27112	0,67575	0,05313	43,30%	56,70%	0,43324	0,46767	0,07263
		Alelo 1	Alelo 2	-					
A27	654	(G) 0,35286	(A) 0,64714	-	47,70%	52,30%	0,47714	0,45735	0,48182
A01	686	(C) 0,63636	(T) 0,36364	-	47,70%	52,30%	0,47727	0,46347	0,64412
A06	702	(A) 0,47933	(T) 0,52067	-	47,10%	52,90%	0,47059	0,49939	0,31601
A08	654	(C) 0,51982	(T) 0,48018	-	49,70%	50,30%	0,49706	0,50000	0,91456
A10	660	(C) 0,64643	(T) 0,35357	-	40,10%	59,90%	0,40061	0,45579	0,03640
A12	660	(G) 0,29913	(A) 0,70087	-	46,40%	53,60%	0,46356	0,41968	0,05249
A14	712	(A) 0,63818	(T) 0,36182	-	43,30%	56,70%	0,43305	0,46247	0,24945
A49	718	(G) 0,47951	(A) 0,52049	-	45,00%	55,00%	0,44954	0,49957	0,07474
A50	706	(G) 0,58522	(A) 0,41478	-	44,50%	55,50%	0,44545	0,48685	0,14320
A51	710	(G) 0,45181	(A) 0,54819	-	43,90%	56,10%	0,43939	0,49575	0,04643
A04	716	(A) 0,52949	(T) 0,47051	-	43,00%	57,00%	0,42978	0,49896	0,00986
A05	700	(G) 0,58275	(A) 0,41725	-	47,90%	52,10%	0,47911	0,48717	0,82849
A37	704	(C) 0,53672	(T) 0,46328	-	54,40%	45,60%	0,54391	0,49820	0,08641
A35	646	(A) 0,70563	(G) 0,29437	-	40,80%	59,20%	0,40845	0,41602	0,79899
A29	680	(G) 0,60446	(A) 0,39554	-	48,00%	52,00%	0,48045	0,47931	1,00000
A36	710	(A) 0,31268	(T) 0,68732	-	49,00%	51,00%	0,49014	0,43043	0,00940
A41	710	(T) 0,49859	(C) 0,50141	-	49,00%	51,00%	0,49014	0,50070	0,74998
A38	714	(G) 0,65966	(C) 0,34034	-	42,30%	57,70%	0,42297	0,44964	0,28788
A16	708	(G) 0,78291	(T) 0,21709	-	35,30%	64,70%	0,35311	0,34248	0,63955
A13	710	(G) 0,70968	(A) 0,29032	-	42,80%	57,20%	0,42817	0,41250	0,52091

As frequências alélicas calculadas para os 30 marcadores InDel em estudo encontram-se representadas graficamente no Gráfico 3.3.3. Novamente, pode-se observar um desbalanceamento entre as frequências alélicas para cada *locus*, com destaque para o HLD70, cuja frequência do alelo para a deleção é de 0,160 (mínima) e do alelo para a inserção de 0,842 (máxima) e para o *locus* HLD58, cujo alelo para a deleção apresenta uma frequência de 0,823 e o alelo para a inserção uma frequência de 0,177. Estas frequências podem ser explicadas pelos mesmos motivos referidos para a população guineense em estudo, isto é, pela ocorrência de *drop out* dos alelos D70- e D58+, apesar de não ter sido observado qualquer desbalanceamento para o *locus* HLD58. No *locus* HLD70, não foi possível determinar a presença do alelo para a deleção num indivíduo apenas, devido ao tamanho muito reduzido e desbalanceado do pico observado, pelo que se pode assumir a ocorrência de *drop out* noutros indivíduos. Também em ambos pode-se considerar estas frequências características da população em estudo. Existem outros *loci* onde o desbalanceamento entre as frequências é notório, mas não tão elevado. Nestes casos, a frequência do alelo menos comum é cerca de 3 vezes menor do que a frequência do alelo mais comum. Contudo, nem todos apresentam grandes diferenças entre os alelos possíveis, como é o caso do *locus* HLD81, HLD133 e HLD39, cujas frequências alélicas se encontram muito próximas dos 0,5.

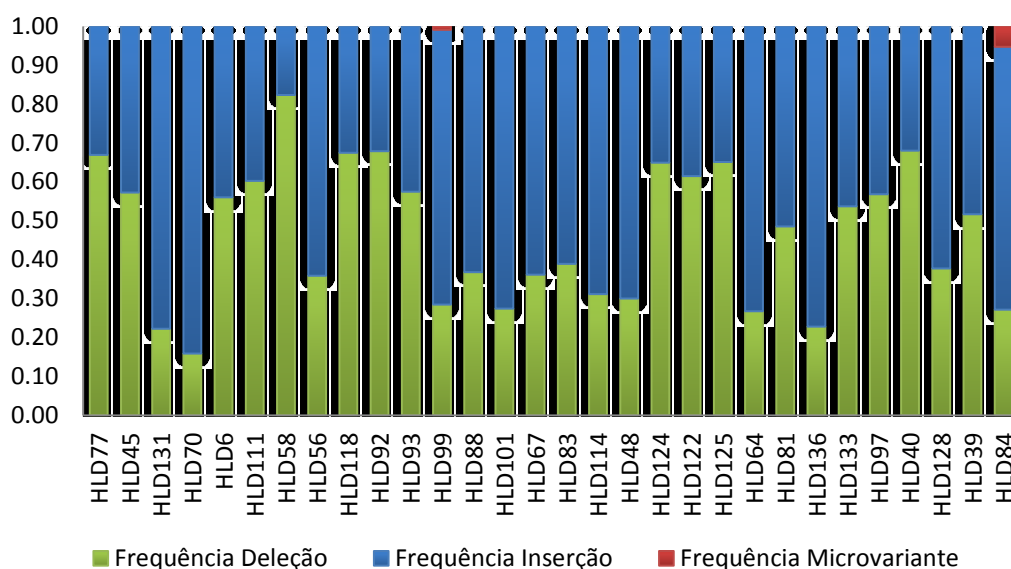


Gráfico 3.3.3 – Frequências alélicas para os 30 marcadores InDel em estudo na população cabo-verdiana imigrante na região de Lisboa.

Relativamente às microvariantes, as frequências para estes dois alelos são muito menores do que as dos restantes, como se pode observar a vermelho no Gráfico 3.3.3. O alelo D99+(-1) foi detetado sempre em heterozigotia em 7 indivíduos, com uma frequência de 0,010. Apesar de a frequência ser menor que 1%, encontram-se muito próxima, podendo chegar a pelo menos 1% com o aumento do número de indivíduos estudados para a população. Já o alelo D84+(-4) foi detetado em 37 indivíduos, dos quais 2 em homozigotia, apresentando uma frequência alélica de 0,053, ou seja, em cerca de 5% da população.

As frequências alélicas para os 20 marcadores SNP estudados (representadas no Gráfico 3.3.4) na população de origem cabo-verdiana, apresentam poucas diferenças entre os dois alelos possíveis, sendo muito próximas dos 0,5 nos *loci* A06, A08, A49, A04 e A41. Os *loci* A12 e A35 apresentam frequências iguais, 0,300 para o alelo constituído por uma guanina e 0,700 para o alelo constituído por uma adenina. Já os *loci* A16 e A13 apresentam uma maior diferença entre a frequência para o alelo 1 e para o alelo 2. O *locus* A16 apresenta para o alelo 1, constituído por uma guanina, uma frequência de 0,783 (máxima) e, para o alelo 2 constituído por uma timina, uma frequência de 0,217 (mínima). O *locus* A13 apresenta para o alelo 1 também constituído por uma guanina, uma frequência de 0,710 e, para o alelo 2 constituído por uma adenina, uma frequência de 0,290. As frequências alélicas detetadas nestes dois *loci* são muito similares às detetadas na população guineense estudada, pelo que se apontam os mesmos motivos para a obtenção destes valores: *drop-out* dos alelos menos comuns e/ou frequências características da população.

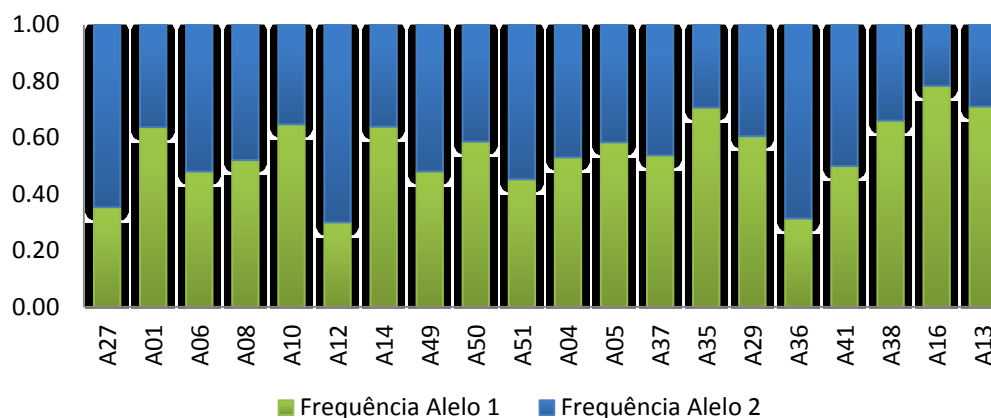


Gráfico 3.3.4 – Frequências alélicas para os 20 marcadores SNP em estudo na população cabo-verdiana imigrante na região de Lisboa. A identificação dos alelos encontra-se na Tabela 3.3.2.

Estes 50 *loci* estudados, apresentam uma frequência de homozigotia média de 55,81% e uma frequência de heterozigotia média de 44,19%, o que significa que existem mais alelos na forma homozigótica do que heterozigótica e que a ocorrência na forma homozigótica é ligeiramente superior e mais provável de ocorrer. Apenas os *loci* HLD131, HLD70 e HLD58 apresentaram níveis de homozigotia (30,2%, 25,7% e 30%, respetivamente) muito inferiores aos níveis de heterozigotia (69,8%, 74,3% e 70%, respetivamente).

A representação gráfica das heterozigotias observadas e esperadas para o total dos 50 marcadores InDel e SNP estudados nesta população pode ser observada no Gráfico 3.3.5. De uma forma geral, as H_o são muito semelhantes às H_e , apresentado valores entre 0,25 e 0,55 e uma média de 0,44 para ambas, com pequenas diferenças apenas, nos *loci* HLD77, HLD6, HLD88, A08 e A29 cuja diferença entre a H_o e a H_e se encontra na ordem das 3 casas decimais. Já os *loci* HLD111, HLD93 HLD67, HLD39, A37 e A36, apresentam maiores diferenças, com uma H_o maior do que a H_e e os *loci* HLD131, HLD83, HLD97, A10, A12, A50, A51, A04 e A38 apresentam uma H_o menor do que a H_e .

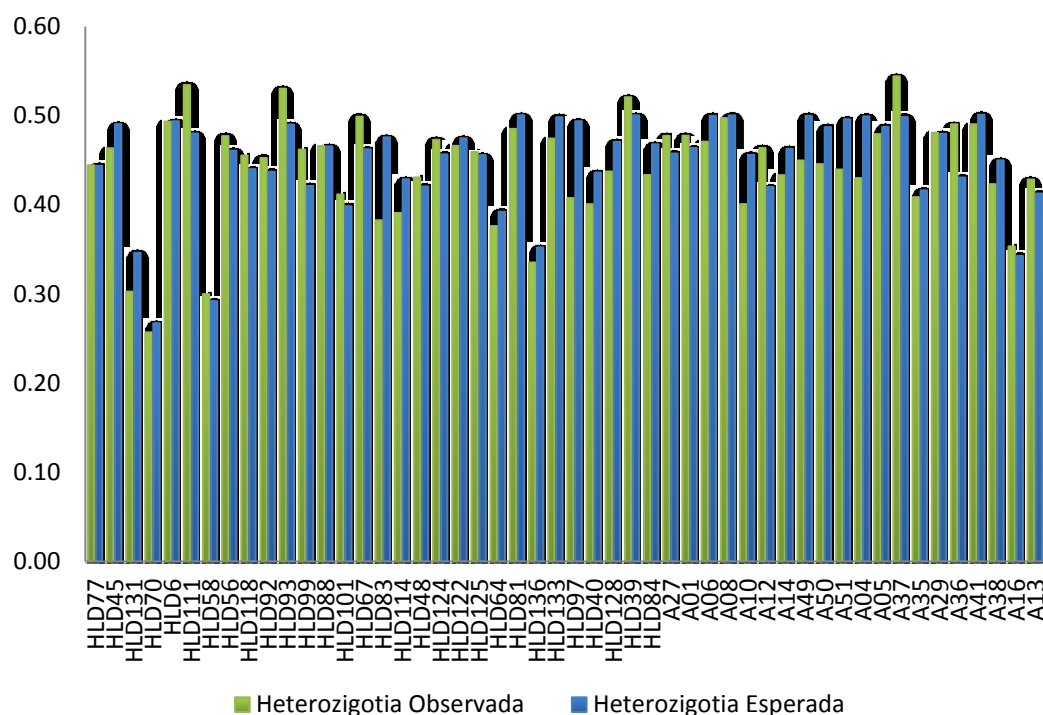


Gráfico 3.3.5 – Heterozigotias observadas e esperadas para os 50 marcadores InDel e SNP estudados na população cabo-verdiana imigrante residente na região de Lisboa.

As diferenças detetadas entre a H_o e a H_e refletem-se no valor de p para o HWE, sendo este menor que 0,05 e estatisticamente significativo nos *loci* HLD131 ($p=0,02326$), HLD111 ($p=0,03779$), HLD83 ($p=0,00017$), HLD97 ($p=0,00096$), A10 ($p=0,03640$), A51 ($p=0,04643$), A4 ($p=0,00986$) e A36 ($p=0,00940$). Após aplicar a correção de Bonferroni ($p<0,001$) apenas os *loci* HLD83 e HLD97 apresentam um desvio do HWE. Os restante *loci* em estudo nesta população encontram-se em HWE ($0,0094<p<1$).

A não detecção de alelos para os *loci* HLD83 (2N=733) e HLD97 (2N=715) pode estar na base do desvio do HWE. No caso do *locus* HLD83 a detecção de um maior número de alelos homozigóticos (61,7%) do que heterozigóticos (38,3%) e a verificação de uma H_o (0,382) menor do que a H_e (0,476) podem explicar a significância obtida para o teste do HWE. No caso do *locus* HLD97, as diferenças entre a frequência de homozigóticos (40,7%) e heterozigóticos (59,3%) e entre a H_o (0,407) e H_e (0,496) são menos significativas, pelo que o desvio do HWE pode ser explicado apenas pelo número reduzido de alelos detetados para a população. Em ambos os casos será necessário um aumento do tamanho da população em estudo, tal como uma otimização da reação de PCR *multiplex*.

3.3.3. Comparação das Populações em Estudo

A comparação entre as duas populações em estudo tem como base os resultados obtidos aquando da caracterização genética com os 50 *loci* de cada população. Numa primeira análise foram comparadas as frequências alélicas obtidas para os 30 *loci* InDel e entre os 20 *loci* SNP, representadas graficamente nos Gráficos 3.3.6 e Gráfico 3.3.7, respetivamente.

Ao observar as frequências alélicas descritas para a população africana (QIAGEN, *SNPforID browser*), verificou-se que o desbalanceamento detetado é característico e esperado para qualquer população de origem de africana, comparativamente à população europeia (QIAGEN, *SNPforID browser*) que apresenta na maioria dos *loci* frequências alélicas próximas de 0,5 (entre 0,327 e 0,673).

Dos *loci* que apresentaram frequências alélicas díspares, 18 revelaram-se comuns entre as duas populações em estudo: HLD131, HLD70, HLD58, HLD118, HLD92, HLD99, HLD88, HLD101, HLD67, HLD83, HLD114, HLD48, HLD124,

HLD125, HLD64, HLD136, A16 e A13. Destes, 9 (HLD131, HLD58, HLD118, HLD92, HLD101, HLD67, HLD48, HLD64 e HLD136) não revelaram qualquer sinal da ocorrência de *drop out* nem de presença de microvariantes que possam afetar a ligação dos *primers* com consequente diminuição da taxa de amplificação ou até mesmo ausência de amplificação. Porém, Kis e colaboradores (2012) verificaram a ocorrência de *drop-out* no *locus* HLD136 e Zidkova e colaboradores (2013) no *locus* HLD58, pelo que este fenómeno pode estar a afetar os resultados obtidos quando a taxa de amplificação é menor, mascarando o desbalanceamento que seria de esperar por comparação com os restantes *loci* quando há ocorrência deste fenómeno. No caso do *locus* A13, o desbalanceamento das frequências alélicas foi também observado na população de Timor-Leste (Santos *et al.*, 2011), indicando uma prevalência do alelo G sobre o alelo A na população africana.

Os restantes *loci* onde as frequências obtidas para os dois alelos constituintes são díspares, isto é, o *locus* HLD128 na população guineense e os *loci* HLD56, HLD122, HLD40, HLD84, A27, A12, A35 e A36 na população cabo-verdiana serão característicos de cada uma das populações em estudo. Isto porque a população com maior número de indivíduos representativos (cabo-verdiana com 367 indivíduos estudados) apresenta um maior número de *loci* nesta condição do que a população guineense com apenas 83 indivíduos, pelo que não será o número de indivíduos representativos a influenciar os resultados obtidos, nestes casos.

Relativamente aos *loci* InDel, as frequências, são de forma geral, similares para a população guineense e cabo-verdiana. Os *loci* HLD45, HLD6, HLD92, HLD122 e HLD84 apresentam frequências muito semelhantes. Os *loci* HLD101, HLD48, HLD125 e HLD128 apresentam maiores diferenças nas frequências alélicas obtidas, sendo a maior diferença entre estas populações verificada no *locus* HLD124, com uma diferença na ordem dos 0,2, sendo a frequência do alelo para deleção maior na população guineense e a frequência do alelo para a inserção maior na população cabo-verdiana. Quanto às microvariantes, as frequências calculadas para o alelo D99+(-1) apresentam uma diferença de 0,02, sendo maior na população guineense (0,036) do que na cabo-verdiana (0,0010). Já as frequências para o alelo D84+(-4) apresentaram-se muito semelhantes, na ordem dos 0,5.

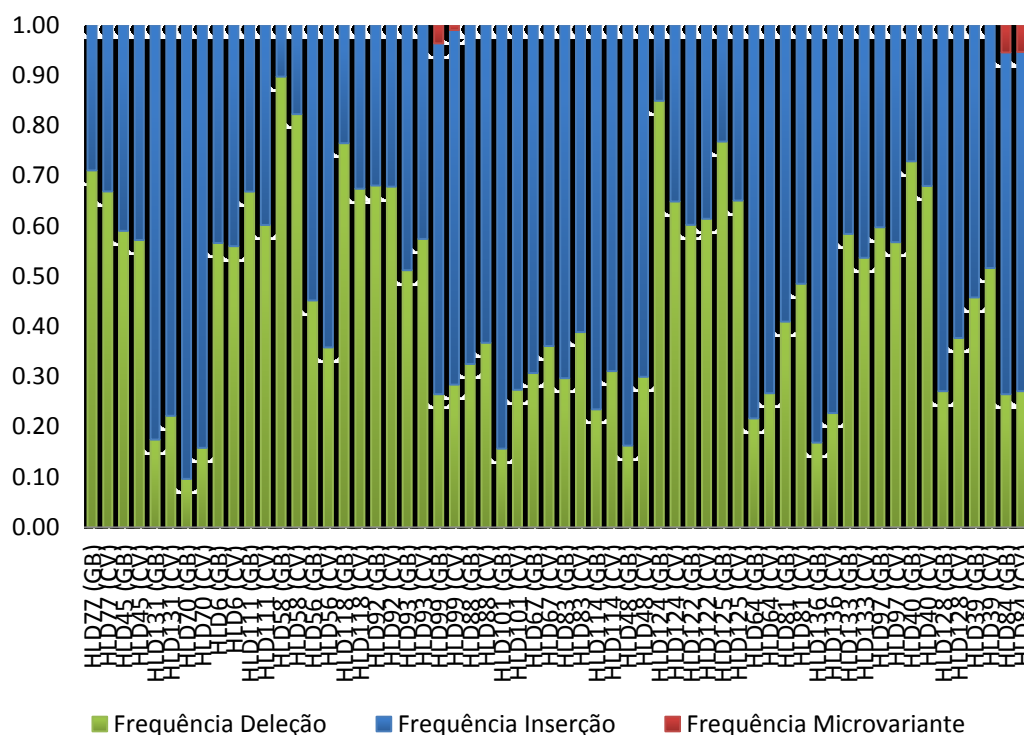


Gráfico 3.3.6 – Frequências alélicas para os *loci* InDel, obtidas para a população imigrante de origem guineense (GB) e de origem cabo-verdiana (CV) residentes na região de Lisboa: comparação entre os resultados obtidos entre as duas populações.

Alguns *loci* SNP estudados, tal como os *loci* InDel, apresentam algumas diferenças nas suas frequências alélicas enquanto outros apresentam frequências muito similares. Os *loci* A05, A16 e A13 apresentam uma maior diferença nas frequências alélicas obtidas, sendo essa de 0,1. Nestes *loci* o alelo 1 correspondente a uma guanina, em ambos, apresenta uma frequência maior na população originária da Guiné-Bissau, enquanto o alelo 2, correspondente a uma adenina, novamente em ambos, apresenta uma frequência maior na população originária de Cabo Verde. O *locus* que apresenta maior semelhança entre estas duas populações é o A51, com uma diferença mínima de 0,003, seguido do *locus* A10 com uma diferença de 0,015.

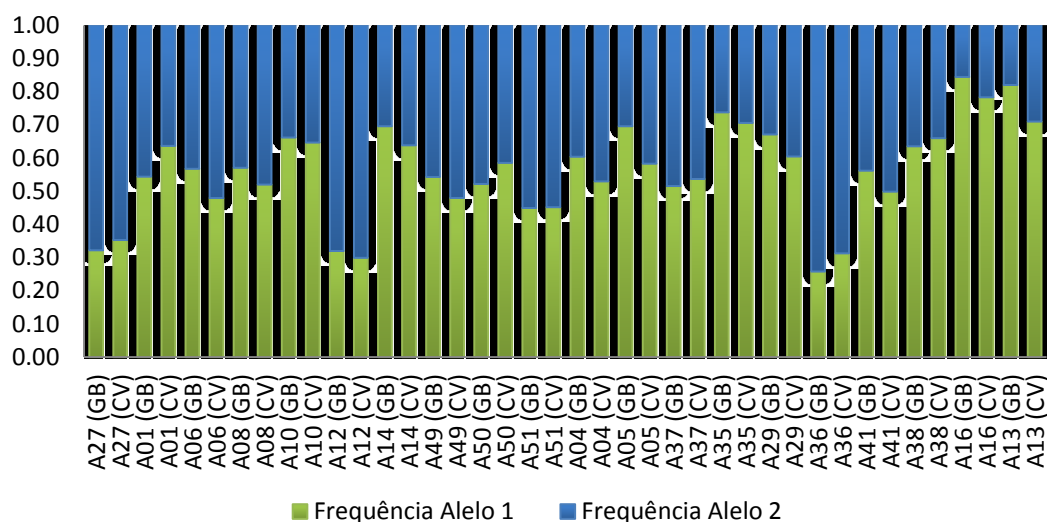


Gráfico 3.3.7 – Frequências alélicas para os *loci* SNP, obtidas para a população migrante de origem guineense (GB) e de origem cabo-verdiana (CV) residentes na região de Lisboa: comparação entre os resultados obtidos entre as duas populações.

Quanto comparadas as H_o , em cada *locus*, entre as duas populações em estudo (Gráfico 3.3.8), verificou-se que a ocorrência de genótipos heterozigóticos é praticamente idêntica apenas nos *loci* HLD56 e A35 e diferente nos *loci* HLD101, HLD124, HLD97, A37 e A29.

O facto de existirem dois *loci*, onde a frequência de heterozigotia é semelhante entre as duas populações, impede o uso destes na determinação da origem geográfica de um indivíduo testado (considerando que exista dúvida entre estas populações apenas), pois a probabilidade de obter um genótipo heterozigótico será também semelhante. Apenas os *loci* onde se verificou maior diferença podem ser assim utilizados quando se pretende determinar a origem geográfica de um qualquer indivíduo testado.

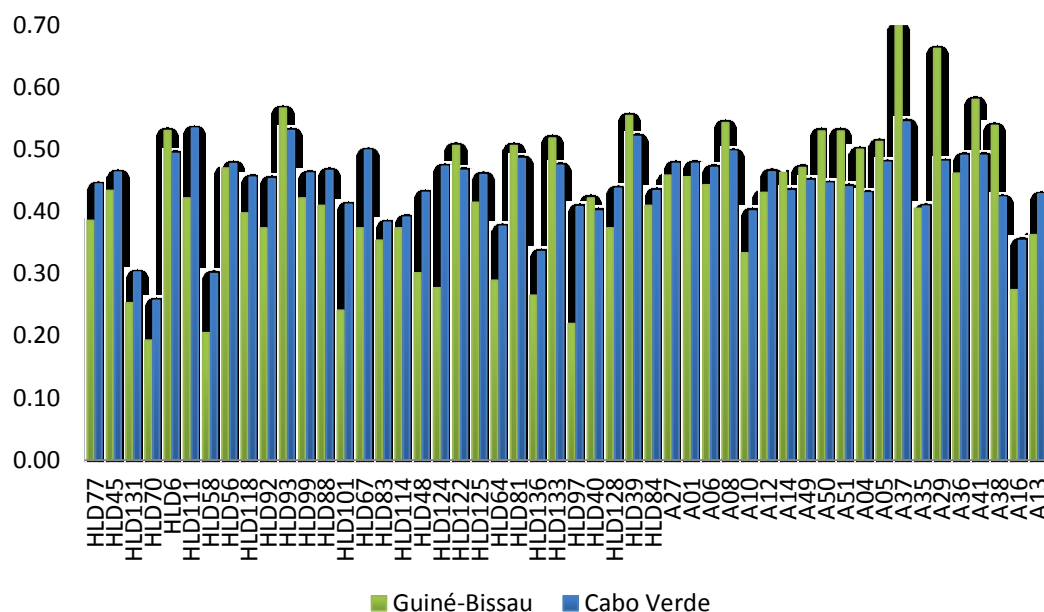


Gráfico 3.3.8 – Heterozigotia Observadas para os 50 loci em estudo na população imigrante de origem guineense e na população imigrante de origem cabo-verdiana.

As diferenças verificadas, em maior número do que as semelhanças, entre estas duas populações, indicam-nos desde já a existência de alguma variabilidade genética entre os indivíduos de origem guineense e os indivíduos de origem cabo-verdiana a residir na região de Lisboa.

3.4. Análise da Variação Intrapopulacional e Interpopulacional

O estudo da variação molecular entre a população imigrante originária da Guiné-Bissau, a população imigrante originária de Cabo Verde e a população Portuguesa, foi realizado de forma a verificar de que forma esta ocorre. As três populações estudadas foram separadas em três grupos: Guiné-Bissau e Cabo Verde, Guiné-Bissau e Portugal e Cabo Verde e Portugal, de forma a verificar as diferenças e semelhanças entre cada população. Os resultados encontram-se resumidos na Tabela 3.4. No anexo 6.1 pode-se verificar os loci onde foram observados a variação máxima e mínima e os loci onde se verificou maior e menor diferenciação genética (F_{st}).

Tabela 3.4.1 – Tabela AMOVA obtida pela análise dos 50 *loci* para 3 grupos de populações: Guiné-Bissau (GB) e Cabo Verde (CV), Guiné-Bissau e Portugal (PT) e Cabo Verde (CV) e Portugal (PT).

	Fonte Variação	Entre populações	Entre indivíduos dentro populações	Entre indivíduos	Total
GB vs CV	Componentes variação	0,08465	0,12214	10,95535	11,16214
	% Variação	0,75835	1,09420	98,14745	
GB vs PT	Componentes variação	105,965	-0,10140	11,52859	12,48684
	% Variação	8,48613	-0,81207	92,32594	
CV vs PT	Componentes variação	0,56368	0,12677	11,32802	12,01847
	% Variação	4,69013	1,05476	94,25511	

De forma geral, nos três grupos populacionais estudados, verificou-se uma maior variação entre os indivíduos (na ordem dos 90%) do que entre as populações, (onde apresenta-se menor que 10%), indicando que é a componente individual da variação genética que afeta a diversidade genética destes *loci*. Estes resultados estão em concordância com os descritos por diversos autores (Bastos-Rodrigues, 2006; Martín *et al.*, 2013; Pepinski *et al.*, 2013; Pereira *et al.*, 2009; Zaumsegel *et al.*, 2013).

A maior percentagem de variação foi verificada quando se comparou a população em estudo da Guiné-Bissau com a população de Portugal (cerca de 8%), sendo a variação entre a população em estudo de Cabo Verde e de Portugal cerca de metade dessa. Entre as populações da Guiné-Bissau e Cabo Verde a percentagem de variação foi menor que 1%. Estes valores são confirmados pela análise dos valores de F_{st} para cada *locus* nos três grupos populacionais definidos (apresentados no anexo 6.2).

Relativamente ao primeiro grupo (Guiné-Bissau e Cabo Verde) verificaram-se diferenças significativas ($p < 0,05$) nos *loci* HLD58, HLD56, HLD118, HLD101, HLD48, HLD124, HLD125, HLD128, A01, A13 e A05. Depois de aplicada a correção de Bonferroni ($p < 0,025$), os *loci* HLD101, HLD48, HLD124, HLD125, HLD128, A13 e A05 continuaram a apresentar diferenças significativas entre as duas populações em estudo, indicando uma maior diferenciação genética e variação devida a estes *loci*. Já os *loci* HLD45, HLD6, HLD92, HLD99, HLD122, HLD84,

A27, A51, A10, A12 e A37 não revelaram qualquer diferença entre as populações em estudo ($p=1,00$), em concordância com os valores de F_{st} negativos. Nos restantes *loci* estudados, verificaram-se valores de F_{st} positivos, porém não significativos ($0,025 < p < 1,000$), pelo que se considera a ocorrência de pequenas diferenças entre as duas populações, porém não úteis quando se pretende uma distinção e uma determinação da origem geográfica de um qualquer indivíduo testado.

Quando se comparou a população da Guiné-Bissau com a população de Portugal, verificou-se a presença de diferenças significativas ($p < 0,05$) em 35 dos 50 *loci*. Desses, após aplicação da correção de Bonferroni ($p < 0,025$), 34 mantiveram-se significativos (apenas o *locus* HLD133 foi corrigido), indicando assim a elevada variação detetada entre estas duas populações. Apenas os *loci* HLD6, HLD93, HLD122, A12 e A37 não revelaram quaisquer diferenças ($p=1,00$). Quanto à comparação da população de Cabo Verde com a população portuguesa, verificaram-se diferenças significativas em 34 dos 50 *loci* estudados. Após a aplicação da correção de Bonferroni ($p < 0,025$) essas diferenças mantiveram-se em 33 dos *loci*. Em 5 *loci* (HLD6, HLD67, HLD122, A06 e A12) não foram detetadas quaisquer diferenças. Em ambos os casos, os restantes *loci* indicaram alguma variabilidade genética, porém não suficiente (não significativa, $0,025 < p < 1,000$) para distinguir as populações determinadas nos grupos.

De forma geral verificaram-se assim maiores diferenças quando se procedeu à comparação das populações em estudo com a população de Portugal do que quando se compararam apenas as populações da Guiné-Bissau e Cabo Verde. Em conjunto, as populações da Guiné-Bissau e de Cabo Verde apresentam 32 *loci* com diferenças significativas da população de Portugal. Isto permite assim uma distinção entre as populações em estudo e a população de Portugal, podendo, após análise mais cuidada, serem utilizados como marcadores AIM de forma a verificar a população de origem de um dado indivíduo testado. Isto será possível porque existem *loci* onde foram verificadas diferenças significativas específicas para os grupos populacionais, nomeadamente o HLD118 e o HLD39 para a população da Guiné-Bissau e Portugal e o A50 para a população de Cabo Verde e Portugal.

Por definição um AIM é um marcador cujas frequências alélicas desbalanceadas e consequentemente baixa heterozigocidade, são características de uma população, permitindo a dedução da origem geográfica, principalmente

continental, de um indivíduo testado do qual não há qualquer tipo de informação. Para além disso um AIM deverá apresentar diferenças significativas entre populações, pois se num *locus* não existir diferenciação entre pelo menos duas populações, significa que essas mesmas são iguais e que a ocorrência de um alelo/genótipo será semelhante e indistinguível (Frudakis *et al.*, 2003; Phillips *et al.*, 2007). Dos 50 *loci*, foram selecionados 9 como possíveis marcadores AIM para a população Africana (Guiné-Bissau e Cabo Verde), por apresentarem diferenças significativas e valores de F_{st} superiores a 0,1 entre a população guineense e portuguesa e entre a população cabo-verdiana e portuguesa: HLD131, HLD70, HLD58, HLD101, HLD114, HLD124, HLD136, A16 e A13. Para a diferenciação da população guineense e da cabo-verdiana não é possível identificar nenhum *locus* com as características mencionadas anteriormente, visto que a semelhança entre estas duas populações é relativamente grande. É ainda possível selecionar os *loci* HLD48, HLD125, HLD64, HLD40, HLD128 e A05 como marcadores AIM para a população da Guiné-Bissau, quando comparada com a de Portugal. Porém é necessário um estudo mais detalhado destes *loci* nos diferentes grupos étnicos da Guiné-Bissau e assumindo a ocorrência de mistura e de subestrutura das populações. Também estudos de ancestralidade com estes marcadores entre estas três populações e outras populações, nomeadamente africanas e europeias serão necessários, de forma a clarificar se os *loci* referidos podem ser utilizados de forma a prever a origem biogeográfica de um qualquer indivíduo.

3.5. Análise Filogenética

A análise filogenética incluiu em primeiro lugar a construção de uma matriz de distâncias genéticas, seguida da construção da árvore filogenética para comparação das duas populações em estudo com sete outras populações estudadas para os mesmos marcadores genéticos.

A construção da matriz de distâncias teve como base as frequências alélicas para cada um dos 50 *loci* estudados. As distâncias genéticas foram calculadas pelos índices de fixação (F_{st}) entre as 9 populações comparadas, encontrando-se na Tabela 3.5.1 a matriz de distâncias global para os 50 *loci* em estudo.

Quanto à população da Guiné-Bissau, verificou-se uma grande semelhança com a população estudada oriunda de Cabo Verde, com uma distância (pequena,

comparativamente com as restantes) apenas de 0,024. Verificou-se também uma proximidade desta população com a população da Somália (0,051), seguida da população de Espanha (0,095) e da população de Portugal (0,100) e verificou-se uma maior diferença com a população da República da China (0,180).

Tabela 3.5.1 – Matriz de distâncias genéticas F_{st} entre as duas populações em estudo (Guiné-Bissau e Cabo Verde) e outras seis populações caracterizadas com os mesmos marcadores genéticos (30 InDel e 20 SNP).

	Guiné-Bissau	Cabo Verde	Portugal	Espanha	Dinamarca	Somália	Alemanha	República da China	Nordeste Itália
Guiné-Bissau	0,000	*	*	*	*	*	*	*	*
Cabo Verde	0,024	0,000	*	*	*	*	*	*	*
Portugal	0,100	0,062	0,000	*	*	*	*	*	*
Espanha	0,095	0,061	0,021	0,000	*	*	*	*	*
Dinamarca	0,121	0,083	0,025	0,031	0,000	*	*	*	*
Somália	0,051	0,041	0,067	0,068	0,092	0,000	*	*	*
Alemanha	0,110	0,073	0,022	0,024	0,025	0,084	0,000	*	*
República da China	0,180	0,144	0,107	0,115	0,112	0,122	0,105	0,000	*
Nordeste Itália	0,098	0,063	0,024	0,026	0,031	0,071	0,025	0,091	0,000

Relativamente à população de Cabo Verde, as frequências alélicas estudadas demonstraram serem mais semelhantes com a Guiné-Bissau como já referido, seguida da população da Somália (0,041), Espanha (0,061) e Portugal (0,062). Novamente verifica-se uma maior diferença quando comparada com a população da República da China (0,144).

Quanto às restantes populações comparadas, verificou-se uma proximidade entre as populações Europeias (Portugal, Espanha, Dinamarca, Alemanha e Nordeste de Itália) e de forma geral, um maior afastamento com a população da República da China (a única população Asiática disponível para comparação).

Estas semelhanças/diferenças podem ser observadas mais claramente na árvore filogenética construída a partir das distâncias calculadas (Figura 3.5.1). Nesta

pode-se observar a existências de três *clusters*, como era de esperar: um para as populações de origem africana, como a Somália, Cabo Verde e Guiné-Bissau, outro para as populações de origem europeia, como Espanha, Portugal, Alemanha, Dinamarca e Nordeste de Itália e outro para a única população representativa da Ásia, a população da República da China, que como já tinha sido observado é a que se encontra mais afastada.

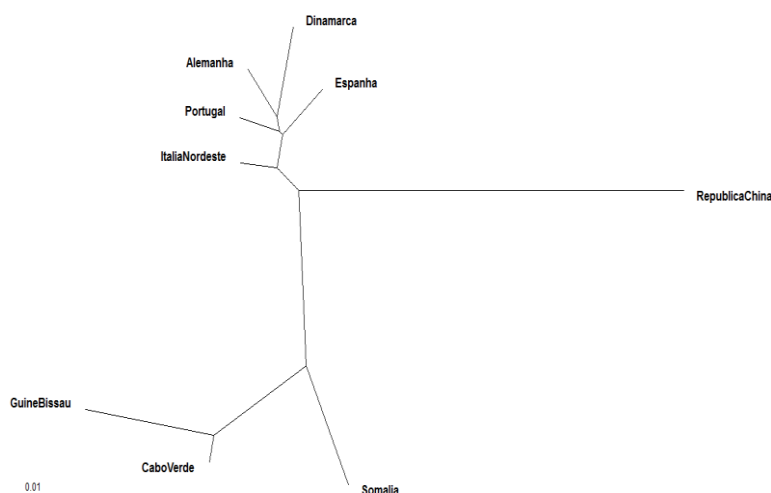


Figura 3.5.1 – Árvore filogenética construída a partir das distâncias genéticas (Fst) demonstradas na Tabela 3.5.1

3.6. Análise de *Linkage Disequilibrium*

Uma das análises necessárias para avaliar o possível uso deste conjunto de marcadores para fins forenses é o teste do LD. Este teste foi realizado numa primeira fase do trabalho pela comparação da localização genómica de cada marcador, que demonstrou uma proximidade menor que 10Mb (Fondevila, *et al.*, 2011) entre os *loci* HLD131e A41. Os dados relativos a estas distâncias genómicas encontram-se no anexo 6.3.

Após a tipagem dos indivíduos em estudo, procedeu-se então ao teste do LD através de todas as combinações possíveis entre os 50 *loci* estudados. Os resultados estão no anexo 6.4. Do total de 1 225 comparações realizadas, verificou-se a existência de LD ($p < 0,05$) em 256 pares de marcadores. Após correção de Bonferroni para testes múltiplos, verificou-se presença de LD ($p < 0,00004$) em 191 pares de *loci*. Destes, verificou-se LD entre dois InDel, HLD99 e HLD124 que se

localizam em cromossomas diferentes (cromossoma 14 e 21, respetivamente) e entre um SNP e um InDel, A01 e HLD83, também em cromossomas diferentes (cromossoma 1 e 8, respetivamente). Os restantes 189 pares de *loci* em LD foram verificados entre SNP ($p=0,00000$), e referem-se a todas as combinações possíveis entre estes *loci*, com exceção do par A10-A36 ($p=0,0001$). O único par onde era prevista a ocorrência de LD seria o HLD131-A41, porém, através deste teste não revelou resultados significativos ($p=0,96604$).

Apesar da extensão de LD poder ser diferente consoante a população em estudo (Kauppi, 2003), apenas os pares A01-A51 e A12-A27 se localizam no mesmo cromossoma (1 e 12, respetivamente), porém em braços cromossómicos diferentes, pelo que apesar dos resultados obtidos não se considera a ocorrência de LD entre os diversos *loci*. O marcador A41 (SNP) foi excluído da análise dos parâmetros forenses pela proximidade física do marcador HLD131 (InDel), sendo necessária uma análise mais cuidada e detalhada deste par, nomeadamente da taxa de recombinação no cromossoma 7.

Esta análise foi realizada para a totalidade dos indivíduos estudados, sem separação por populações, pelo que pode haver um fenómeno de subestruturação a afetar a significância de LD. Outros fatores populacionais podem estar também a influenciar os resultados obtidos, como o tamanho da população em estudo e os fenómenos de seleção e deriva genética (Ardlie *et al.*, 2002; Bhangale *et al.*, 2005; Brookes, 1999; Cardon e Bell, 2001; Kauppi, 2003). LaRue e colaboradores (2012) sugerem que 8 pares de InDel, apesar de não apresentarem resultados significativos de LD, podem estar em *linkage* com taxas de recombinação menores que 0,5 por se localizarem no mesmo cromossoma, separados por menos de 50Mb. Outros autores, ao estudarem os mesmos 30 InDel e 20 SNP não verificaram qualquer indício da ocorrência de LD entre os *loci* InDel e os *loci* SNP (da Silva *et al.*, 2013; Kis *et al.*, 2012; Montelius *et al.*, 2009; Pepinski *et al.*, 2013; Pereira *et al.*, 2008; Romanini *et al.*, 2012; Sanchez *et al.*, 2006; Seong *et al.*, 2014; Farzad *et al.*, 2013; Tomas *et al.*, 2013; Vullo *et al.*, 2011; Wei *et al.*, 2014; Zidkova *et al.*, 2013). Apenas Martín e colaboradores (2013) observaram uma pequena evidência de LD entre pares de *loci* InDel, porém insignificante após aplicarem a correção de Bonferroni, pelo que consideram também os 30 *loci* independentes. Já Poulsen e colaboradores (2011)

verificaram a existência de LD apenas nos pares A05-A38 e A05-A08, como verificado neste estudo.

O facto de os resultados observados serem contraditórios aos observados pelos autores referidos e tendo em conta a falta de estudos acerca destes marcadores nas populações estudadas, o efeito dos fatores populacionais é assim a única explicação para o elevado número de pares de marcadores em LD.

3.7. Parâmetros Forenses

Para utilizar um conjunto de marcadores genéticos em investigações forenses é necessária uma avaliação estatística de cada marcador e do conjunto, de forma a garantir a correta individualização de um qualquer indivíduo testado.

Os parâmetros forenses foram calculados para todos os 50 marcadores genéticos (30 InDel e 20 SNP) e para ambas as populações em estudo. Os valores obtidos para a PM, o PD, o PIC, o PE e o TPI de cada marcador, nas populações estudadas encontram-se na Tabela 3.7.1.

Tabela 3.7.1 – Parâmetros forenses calculados para os 50 marcadores genéticos estudados na população guineense e cabo-verdiana imigrante a residir na região de Lisboa. Os valores máximos e mínimos obtidos encontram-se destacados.

<i>Locus</i>	PM		PD		PIC		PE		TPI	
	Guiné-Bissau	Cabo Verde	Guiné-Bissau	Cabo Verde	Guiné-Bissau	Cabo Verde	Guiné-Bissau	Cabo Verde	Guiné-Bissau	Cabo Verde
HLD77	0,426	0,409	0,574	0,591	0,33	0,34	0,105	0,143	0,81	0,90
HLD45	0,365	0,369	0,635	0,631	0,37	0,37	0,136	0,880	0,88	0,93
HLD131	0,555	0,489	0,445	0,511	0,25	0,29	0,046	0,065	0,67	0,72
HLD70	0,689	0,575	0,311	0,425	0,16	0,23	0,028	0,047	0,62	0,67
HLD6	0,400	0,379	0,600	0,621	0,37	0,37	0,215	0,182	1,06	0,99
HLD111	0,402	0,415	0,598	0,585	0,34	0,36	0,128	0,219	0,86	1,07
HLD58	0,674	0,544	0,326	0,456	0,17	0,25	0,031	0,063	0,63	0,71
HLD56	0,366	0,404	0,634	0,596	0,37	0,35	0,162	0,168	0,94	0,96
HLD118	0,480	0,416	0,520	0,584	0,29	0,34	0,112	0,151	0,83	0,92
HLD92	0,401	0,418	0,599	0,582	0,34	0,34	0,099	0,149	0,80	0,91
HLD93	0,415	0,402	0,585	0,598	0,37	0,37	0,252	0,215	1,15	1,06
HLD99	0,386	0,428	0,614	0,572	0,37	0,34	0,128	0,156	0,86	0,93
HLD88	0,403	0,395	0,597	0,605	0,34	0,36	0,120	0,159	0,85	0,94
HLD101	0,582	0,445	0,418	0,555	0,23	0,32	0,042	0,121	0,66	0,85
HLD67	0,410	0,413	0,590	0,587	0,34	0,35	0,099	0,186	0,80	1,00
HLD83	0,415	0,362	0,585	0,638	0,33	0,36	0,088	0,104	0,77	0,81

Tabela 3.7.1 – Continuação - Parâmetros forenses calculados para os 50 marcadores genéticos estudados na população guineense e cabo-verdiana imigrante a residir na região de Lisboa. Os valores máximos e mínimos obtidos encontram-se destacados.

<i>Locus</i>	PM		PD		PIC		PE		TPI	
	Guiné-Bissau	Cabo Verde	Guiné-Bissau	Cabo Verde	Guiné-Bissau	Cabo Verde	Guiné-Bissau	Cabo Verde	Guiné-Bissau	Cabo Verde
HLD114	0,476	0,410	0,524	0,590	0,29	0,34	0,099	0,108	0,80	0,82
HLD48	0,562	0,428	0,438	0,572	0,24	0,33	0,064	0,134	0,72	0,88
HLD124	0,582	0,407	0,418	0,593	0,22	0,35	0,055	0,165	0,69	0,95
HLD122	0,399	0,386	0,601	0,614	0,36	0,36	0,193	0,159	1,01	0,94
HLD125	0,487	0,402	0,513	0,598	0,29	0,35	0,123	0,154	0,85	0,92
HLD64	0,497	0,445	0,503	0,555	0,28	0,31	0,059	0,100	0,70	0,80
HLD81	0,394	0,368	0,606	0,632	0,37	0,37	0,193	0,175	1,01	0,97
HLD136	0,560	0,482	0,440	0,518	0,24	0,29	0,050	0,079	0,68	0,75
HLD133	0,399	0,366	0,601	0,634	0,37	0,37	0,204	0,166	1,04	0,95
HLD97	0,372	0,349	0,628	0,651	0,37	0,37	0,035	0,118	0,64	0,84
HLD40	0,450	0,405	0,550	0,595	0,32	0,34	0,128	0,114	0,86	0,83
HLD128	0,441	0,380	0,559	0,620	0,32	0,36	0,099	0,138	0,80	0,89
HLD39	0,410	0,386	0,590	0,614	0,37	0,37	0,239	0,206	1,12	1,04
HLD84	0,353	0,355	0,647	0,645	0,40	0,40	0,120	0,135	0,85	0,88
A27	0,417	0,408	0,583	0,592	0,34	0,35	0,154	0,168	0,92	0,96
A01	0,360	0,402	0,640	0,598	0,37	0,36	0,152	0,168	0,92	0,96
A06	0,360	0,363	0,640	0,637	0,37	0,37	0,142	0,163	0,90	0,94
A08	0,409	0,374	0,591	0,626	0,37	0,37	0,228	0,185	1,09	0,99
A10	0,389	0,385	0,611	0,615	0,35	0,35	0,078	0,114	0,75	0,83
A12	0,418	0,440	0,582	0,560	0,34	0,33	0,134	0,158	0,88	0,93
A14	0,433	0,386	0,567	0,614	0,33	0,36	0,157	0,135	0,93	0,88
A49	0,363	0,355	0,637	0,645	0,37	0,37	0,163	0,147	0,94	0,91
A50	0,391	0,366	0,609	0,634	0,37	0,37	0,215	0,144	1,06	0,90
A51	0,398	0,355	0,602	0,645	0,37	0,37	0,215	0,140	1,06	0,89
A04	0,399	0,349	0,601	0,651	0,36	0,37	0,188	0,133	1,00	0,88
A05	0,455	0,379	0,545	0,621	0,33	0,37	0,193	0,170	1,01	0,96
A37	0,535	0,402	0,465	0,598	0,37	0,37	0,428	0,229	1,67	1,10
A35	0,451	0,426	0,549	0,574	0,31	0,33	0,117	0,119	0,84	0,85
A29	0,553	0,387	0,447	0,613	0,34	0,36	0,372	0,171	1,48	0,96
A36	0,477	0,440	0,523	0,560	0,31	0,34	0,156	0,179	0,93	0,98
A41	0,432	0,370	0,568	0,630	0,37	0,37	0,268	0,179	1,19	0,98
A38	0,429	0,396	0,571	0,604	0,36	0,35	0,223	0,129	1,08	0,87
A16	0,564	0,492	0,436	0,508	0,24	0,28	0,053	0,088	0,69	0,77
A13	0,538	0,435	0,462	0,565	0,25	0,33	0,093	0,132	0,78	0,87

Para ambas as populações imigrantes residentes na região de Lisboa estudadas, isto é, com origem na Guiné-Bissau e com origem em Cabo Verde, as PM calculadas encontram-se entre 0,353 (HLD84) e 0,689 (HLD70) e entre 0,349 (HLD97 e A04) e 0,575 (HLD 70), respetivamente. Quanto ao PD, este apresenta um valor mínimo de 0,311 e 0,425 (HLD70) e um valor máximo de 0,647 (HLD84) e de 0,651 (HLD97 e A04) na população guineense e cabo-verdiana, respetivamente. Os valores de PIC calculados apresentam por sua vez, um mínimo de 0,160 (Guiné-Bissau) e 0,230 (Cabo Verde) para o *locus* HLD70 e 0,400 (em ambas as populações em estudo) para o *locus* HLD84. Estes valores apesar de indicarem um grau de polimorfismo baixo, são os esperados para marcadores bialélicos, que nunca apresentarão um PIC superior a 0,375 (Hildebrand *et al.*, 1992) Relativamente ao PE, apresenta um mínimo de 0,028 (Guiné-Bissau) e 0,047 (Cabo-Verde) no *locus* HLD70 e um máximo de 0,428 (Guiné-Bissau) e 0,880 (Cabo Verde) para os *loci* A37 e HLD45, respetivamente. Por último, os valores de TPI obtidos apresentam um mínimo de 0,620 (Guiné-Bissau) e de 0,670 (Cabo Verde) para o *locus* HLD70 e um máximo de 1,670 (Guiné-Bissau) e de 1,100 (Cabo Verde) para o *locus* A37.

Tendo em conta que estes marcadores são bialélicos, os resultados obtidos para os diversos parâmetros são satisfatórios para serem utilizados em investigações forenses, com exceção do *locus* HLD70 que revelou-se ser pouco informativo, apresentando os valores mínimos obtidos para ambas as populações em 4 dos 5 parâmetros calculados e indicando uma elevada probabilidade de obter genótipo idênticos em dois indivíduos selecionados ao acaso (PM). Também a deteção de um terceiro alelo nos *loci* HLD99 e HLD84 permitiu aumentar um pouco o poder de discriminação destes marcadores. Estes resultados estão de acordo com os descritos em estudos realizados para os 30 *loci* InDel (Akhteruzzaman *et al.*, 2013; Carvalho e Pinheiro, 2013; Fondevila *et al.*, 2012; Friis *et al.*, 2012; Kim *et al.*, 2013; Kis *et al.*, 2012; LaRue *et al.*, 2012; Martín *et al.*, 2013; Neuvonen *et al.*, 2012; Pepinski *et al.*, 2013; QIAGEN; Saiz *et al.*, 2013; Seong *et al.*, 2014; da Silva, *et al.*, 2013; Turrina *et al.*, 2011; Zidkova *et al.*, 2013). Relativamente aos 20 marcadores SNP, só é possível a comparação com os estudos realizados pela metodologia 52-plex (onde estão inseridos os 20 SNP), onde também se verificou uma concordância com os resultados obtidos por Barbaro e colaboradores (2012) na população de Italia e por Santos e colaboradores (2011) na população de Timor-Leste. As pequenas variações

observadas devem-se apenas ao diferente número de indivíduos estudados (quanto maior, mais precisos são os resultados) e às diferentes populações em cada estudo.

Os parâmetros forenses foram também calculados para a totalidade dos marcadores, isto é, para 49 marcadores. Os valores obtidos da probabilidade de *matching* combinada (PMC), para o poder de discriminação combinado (PDC), para o poder de exclusão combinado (PEC) e para o índice de paternidade típico combinado (TPIC) encontram-se na Tabela 3.7.2 para ambas as populações em estudo. Estes mesmos parâmetros, quando calculados separadamente para os 19 marcadores SNP revelaram-se, como seria de esperar (Dario *et al.*, 2009), baixos. Os 49 marcadores, quando utilizados em conjunto, revelaram-se bastante informativos para serem utilizados na rotina forense, com uma PMC muito reduzida e um PDC muito próximo de 100% o que indica que a probabilidade de dois indivíduos selecionados ao acaso terem genótipos iguais é quase nula. Também o PEC e o TPIC são, para ambas as populações, superiores a 99%.

Tabela 3.7.2 – Parâmetros Forenses combinados para os 49 marcadores Genéticos em estudo.

	PMC	PDC	PEC	TPIC
Guiné-Bissau	$5,42 \times 10^{-18}$	>99,99%	99,95%	99,84%
Cabo Verde	$5,96 \times 10^{-20}$	>99,99%	99,99%	99,47%

Sendo a primeira vez que este painel de marcadores genéticos está a ser testado, não existe termo de comparação para os parâmetros forenses combinados. De qualquer forma, como seria de esperar, verificou-se uma PMC superior à observada por diversos autores quando analisaram os 30 *loci* InDel (Akhteruzzaman *et al.*, 2013; Carvalho e Pinheiro, 2013; Silva *et al.*, 2013; Fondevila *et al.*, 2012; Friis *et al.*, 2012; LaRue *et al.*, 2012; Neuvonen *et al.*, 2012; Pepinski *et al.*, 2013; Seong *et al.*, 2014) e à verificada noutros painéis desenvolvidos com marcadores InDel (Li *et al.*, 2011; Pereira *et al.*, 2009b; Pimenta e Pena, 2010). Devido ao número próximo de marcadores bialélicos, a PMC obtida foi semelhante à verificada nos estudos com recurso à metodologia SNPforID 52-plex (Pereira *et al.*, 2008;

Sanchez *et al.*, 2006; Santos *et al.*, 2011) e 49-plex (adaptada da anterior) (Poulsen *et al.*, 2011; Tomas *et al.*, 2013).

3.8. Discussão

Este estudo teve como principal objetivo a caracterização de duas populações imigrantes em Portugal, nomeadamente a população de origem guineense e a população de origem cabo-verdiana, muito representativas na região Sul de Portugal. Essa caracterização foi realizada com recurso a marcadores genéticos forenses menos estudados. Estes podem vir a servir de alternativa ou como complemento aos STR utilizados rotineiramente, nas diversas investigações forenses. Para além disso as ligações históricas que envolvem estas populações e a população de Portugal foram também um motivo para o seu estudo, de forma a tentar compreender as diferenças e semelhanças entre estas três populações.

A caracterização destas populações revelou uma elevada heterogeneidade nas frequências alélicas obtidas, contrariamente ao que seria de esperar para marcadores bialélicos, que deveriam apresentar frequências alélicas próximas ou iguais a 0,5 (Amorim e Pereira, 2005). Esta heterogeneidade é uma mais-valia quando se pretende caracterizar populações e determinar marcadores AIM para determinação da origem geográfica, pois o desbalanceamento das frequências alélicas indica-nos que a ocorrência de um dado alelo numa população é mais frequente do que a ocorrência do outro alelo constituinte do mesmo *locus* bialélico, nessa mesma população. Para a população guineense e cabo-verdiana, isto é, africana, foram indicados 9 possíveis *loci* que podem vir a ser utilizados como AIMs (HLD131, HLD70, HLD58, HLD101, HLD114, HLD124, HLD136, A16 e A13) e considerando apenas a população guineense, 6 *loci* (HLD48, HLD125, HLD64, HLD40, HLD128 e A05). Kim e colaboradores (2013) e Pepinski e colaboradores (2013) também identificaram vários *loci* InDel como marcadores AIM, mas para populações Asiáticas e para as populações da República da China e Polónia, respetivamente. Assim sendo, com base nestes marcadores já é possível a diferenciação de pelo menos duas grandes populações: Africana e Asiática. Relativamente aos *loci* SNP, 3 estão incluídos no painel de marcadores de ancestralidade SNPforID 34-plex (A07, A29 e A13) por permitirem a distinção das populações com origem na África Subsariana, na Europa e na Ásia. (Fondevila *et al.*, 2013; Phillips *et al.*, 2007) Ainda assim, é necessária

uma melhor compreensão acerca da estrutura e da evolução das principais populações de forma determinar os marcadores que as permitem distinguir (Bastos-Rodrigues *et al.*, 2006; Frudakis *et al.*, 2003; Hinds *et al.*, 2005; Nievergelt *et al.*, 2013; Pereira e Gusmão, 2012; Romualdi *et al.*, 2002; Santos *et al.*, 2010; Yang *et al.*, 2005; Zaumsegl *et al.*, 2013), e só então definir marcadores para populações mais específicas, como por exemplo da Guiné-Bissau, de Cabo Verde e de Portugal, que seriam de maior interesse para o SGBF-INMLCF-Sul na resolução de casos criminais. É necessário realçar que tanto o painel de marcadores InDel como o painel de marcadores SNP utilizados não foram desenhados para distinguir populações e origens geográficas, mas sim para identificação individual na população Europeia no caso dos *loci* InDel (QIAGEN, 2013) e no caso dos SNP (Sanchez *et al.*, 2006), pelo que os resultados obtidos revelaram-se favoráveis na possibilidade de utilizar estes painéis de marcadores também para estudos de ancestralidade.

A origem do nativo cabo-verdiano teve como base a miscigenação entre os homens europeus, nomeadamente portugueses, e as escravas negras da costa africana, principalmente da Guiné-Bissau (Barcellos, 1899) que era um local de eleição na captura dos escravos. Os resultados obtidos aquando do cálculo dos valores de *Fst* entre a população guineense e cabo-verdiana demonstraram uma grande semelhança entre ambas, com apenas 7 dos 50 *loci* estudados a apresentarem diferenças significativas e um *Fst* médio de 0,008, confirmando o papel dos indivíduos guineenses na origem do povo cabo-verdiano. Também Brehm e Pereira (2002) e Morais (2013), ao estudarem as linhagens maternas na população cabo-verdiana, verificaram uma elevada contribuição proveniente de mulheres oriundas da costa ocidental de África, para a origem da composição genética das linhagens maternas atuais no país. Quando se calculou os valores de diferenciação genética entre a população de Cabo Verde e a população de Portugal, os resultados não foram tão satisfatórios como se esperava, tendo em conta o grande papel de Portugal na colonização das ilhas. Foram observadas diferenças significativas em 33 dos 50 *loci* estudados, com uma diferenciação média de 0,047, muito superior à observada anteriormente entre as populações em estudo. Esta diferenciação genética indica a fraca contribuição da população portuguesa, comparativamente à da Guiné-Bissau, na constituição genética dos cabo-verdianos. De realçar que 5 *loci* (HLD6, HLD67, HLD122, A06 e A12) revelaram-se idênticos entre estas duas populações, podendo

estes, representar a contribuição da população portuguesa. De encontro a estes resultados, Gonçalves e colaboradores (2003) ao estudarem as linhagens paternas presentes na constituição genética das ilhas verificaram uma baixa contribuição dos homens Europeus nessas linhagens, indicando os homens provenientes do Médio Oriente, nomeadamente Judeus, como principais contribuidores. Também Fernandes e colaboradores (2003) com recurso a marcadores STR e Spínola e colaboradores (2002) com recurso a marcadores HLA (do inglês *Human Leukocyte Antigens*), verificaram que a população de Cabo Verde é geneticamente atípica, comparativamente às restantes populações da zona Subsariana, com marcadores típicos das populações Europeias e marcadores típicos da África Ocidental. Apesar da clara evidência da presença de homens Europeus e de mulheres da região Subsariana como contribuidores das atuais características genéticas da população Cabo Verdiana, não existem evidências claras da esperada grande contribuição da população portuguesa. Para além disso, quando se procedeu ao estudo das relações filogenéticas com recurso às frequências alélicas, verificou-se uma maior proximidade com a população espanhola do que com a população portuguesa, indicando uma maior presença da primeira durante a colonização das ilhas do que a referida na literatura histórica. Serão necessários estudos mais específicos que incluam a população portuguesa e espanhola e as diferentes populações da zona Ocidental de África, de forma a compreender exatamente as contribuições destas na origem e constituição genética da população cabo-verdiana. O uso de marcadores AIM será também uma mais-valia na compreensão da estrutura desta população, quando se estudarem indivíduos cabo-verdianos emigrantes, já estabelecidos na nossa população.

Relativamente ao estudo da população da Guiné-Bissau e de Portugal, os resultados obtidos indicam uma diferenciação genética ainda maior entre estas duas populações com um valor de F_{st} médio de 0,100, como seria de esperar, visto que à data da chegada dos portugueses já existiam diversos grupos étnicos a habitar na região. Também pela observação do F_{st} global para a distribuição das frequências alélicas nestas populações, a distância genética é bem visível e claramente representada na árvore filogenética apresentada Figura 3.5.1 (página 88). Novamente verificou-se uma maior proximidade com a população espanhola, apesar de pouco estar referido acerca do estabelecimento desta nos Rios da Guiné, para além da

captura ilegal de escravos. Tendo em conta a existência de influências europeias nas linhagens paternas, verificadas por Rosa e colaboradores (2007) e Carvalho e colaboradores (2011), a influência dos portugueses e de outras populações europeias não pode ser excluída da actual constituição genética da população da Guiné-Bissau. A constituição genética tipicamente Africana é clara nesta população, como observado pela proximidade ($F_{st} < 0,1$) desta população com a população da Somália, tal como a diferenciação de alguns grupos étnico-linguísticos, verificada através das linhagens maternas (Carvalho *et al.*, 2011; Rosa *et al.*, 2004). Essa diferenciação dos diversos grupos étnicos deverá ser estudada na população guineense imigrante em Portugal, pois diferentes constituições genéticas (por mínimas que sejam) terão diferentes influências na população portuguesa e consequentemente na constituição genética portuguesa, quer seja nas linhagens paternas, como nas maternas e nas regiões autossómicas. Novamente, sugere-se também o uso de marcadores AIM para o estudo da população guineense emigrante em Portugal.

Para entender a proximidade genética verificada nas populações em estudo com a população espanhola será necessária uma análise mais rigorosa, nomeadamente das diferenciações genéticas por *locus*, que poderão elucidar acerca da contribuição genética da população espanhola na constituição genética atual das populações de Cabo Verde e da Guiné-Bissau. Também o aumento do número de indivíduos representativos de cada população será necessário, principalmente na população da Guiné-Bissau, onde foram estudados apenas 83, de forma a aumentar o grau de confiança nos resultados obtidos.

Este estudo teve também como objetivo a utilização de marcadores genéticos recentemente utilizados, nomeadamente marcadores InDel e SNP, de forma a avaliar de forma muito simples a sua utilização em investigações forenses.

Relativamente ao estudo dos marcadores InDel, este foi realizado com recurso ao *kit* Investigator[®] DIPplex (*QIAGEN*), desenhado para identificação individual em populações principalmente Europeias, pelo que pode não ser o mais indicado para estudar indivíduos de origem africana, visto que foram verificados desbalanceamentos heterozigóticos e *drop-out* alélicos, nomeadamente no marcador HLD97 (Fondevila *et al.*, 2011, 2012; LaRue *et al.*, 2012). A deteção do alelo para a inserção *locus* HLD97 revelou-se assim problemática, com consequente influência na análise estatística realizada, principalmente nos valores de heterozigotia

observados (menores do que era esperado), que por sua vez influenciaram o facto de este *locus* se apresentar desviado do HWE nas duas populações estudadas. Também o *locus* HLD83 na população estudada de Cabo Verde encontra-se desviado do HWE, neste caso devido ao grande desbalanceamento heterozigótico verificado. A influência dos fatores populacionais, apesar de não excluída como hipótese, será muito diminuta, pois caso as populações estudadas não fossem ideais ter-se-ia observado mais *loci* com desvios significativos do HWE. Para estes casos, tal como os autores já referidos, sugere-se uma reformulação dos *primers* incluídos no *kit*, de forma a evitar problemas na ligação desses à região de ADN alvo e consequente baixa taxa de amplificação. Contudo nem tudo são desvantagens, visto que, a deteção de duas microvariantes alélicas, presentes em pelo menos 1% nas duas populações estudadas permitiu aumentar o poder de discriminação do painel de 30 marcadores, com dois marcadores analisados (HLD99 e HLD84) sendo trialélicos e não bialélicos.

Quanto ao painel de 20 SNP estudado em 2 reações de PCR *multiplex*, a obtenção de um perfil completo com uma só tipagem revelou-se difícil. Diversos autores mesmo após realizarem a tipagem dos mesmos indivíduos 3 e 4 vezes, obtiveram perfis incompletos (Børsting *et al.*, 2009; Montelius *et al.*, 2009; Sanchez *et al.*, 2006), revelando a necessidade de uma otimização das reações de PCR *multiplex* para o estudo de marcadores SNP antes de serem utilizados em investigações forenses, onde frequentemente a quantidade de amostra biológica é muito reduzida e limitada, impedindo a tipagem da mesma diversas vezes. Também nestes *loci* foram observados fenómenos de *drop-out* e desbalanceamento, nomeadamente nos *loci* A13 e A16, que podem levar a uma determinação incorreta do genótipo de um indivíduo, com consequência no resultado da investigação forense realizada.

Apesar das dificuldades na amplificação e deteção dos *loci* em estudo, quando se analisaram os diversos parâmetros forenses, verificou-se estes que são adequados para uso na área forense, sendo os valores obtidos para os 50 marcadores semelhantes aos valores dos *kits* comerciais de marcadores STR utilizados nos laboratórios de Genética Forense. No entanto é necessária uma especial atenção à possibilidade de alguns marcadores estudados estarem em LD (invalidando a sua utilização em investigações forenses por não poderem ser tratados estatisticamente

como independentes) bem como aquando do estudo destes em conjunto ou como complemento com os marcadores STR. Dois estudos acerca da ocorrência de LD entre marcadores InDel e STR demonstraram apenas um par de marcadores em localização próxima e com uma taxa de recombinação baixa: HLD136-D22S1045 (Fondevila *et al.*, 2011, 2012).

Os resultados obtidos aquando da caracterização de cada população irão permitir a construção de uma base de dados, uma para a população imigrante da Guiné-Bissau e outra para a população imigrante de Cabo Verde a residir na região de Lisboa. Como estas populações representam 4% e 10%, respetivamente, da população total de imigrantes a residir legalmente em Portugal, as bases de dados serão de maior interesse no SGBF-INMLCF quer em investigações de parentesco como criminais, na medida em que sabendo à partida a origem geográfica, o cálculo da frequência de um qualquer perfil de ADN obtido será mais específico.

4. Conclusões

O estudo de 83 indivíduos de origem guineense e de 367 indivíduos de origem cabo-verdiana, a residir legalmente na região Sul de Portugal com um painel de 50 marcadores genéticos (30 InDel e 20 SNP), permitiu formular um capítulo de conclusões com potencial interesse e impacto nas áreas forenses e da genética populacional.

4.1. Estudo dos marcadores InDel e SNP

- A genotipagem dos indivíduos em estudo revelou-se mais trabalhosa do que seria de esperar, com deteção de desbalanceamentos e *drop-outs* em diversos alelos e/ou marcadores, pelo que a utilização deste painel encontra-se limitada para usos forenses, nomeadamente para investigações criminais, quando se está perante indivíduos de origem africana;

- As microvariantes detetadas nos *loci* HLD99 e HLD84, verificadas na população africana, permitem um aumento do poder de discriminação do painel de marcadores estudados, devido de um outro InDel que se localiza dentro da região de amplificação destes *loci*.

- Serão necessárias melhorias nas duas metodologias utilizadas, nomeadamente no desenho dos *primers*, de forma a permitir uma correta obtenção dos perfis genéticos e uma correta análise estatística aquando de uma investigação forense;

- Quando utilizado o painel de 50 marcadores, os parâmetros forenses calculados revelaram-se adequados para fins forenses, sendo esses resultados semelhantes aos dos *kits* de marcadores STR utilizados na rotina forense,

- Verificou-se a ocorrência de LD entre marcadores, contrariamente ao descrito por outros autores.

4.2. Estudo das Populações imigrantes da Guiné-Bissau e de Cabo Verde a residir na região de Lisboa

- A caracterização das populações em estudo foi útil na construção de uma base de dados de frequências alélicas para que aquando da comparência de um qualquer indivíduo de origem guineense ou cabo-verdiana no SGBF-INMLCF-Sul o cálculo da frequência de um genótipo obtido seja mais preciso;

- As diferenças detetadas entre as populações estudadas e a população de Portugal são um ponto de partida para a diferenciação de indivíduos pela sua origem geográfica, utilizando para isso, marcadores que para além de permitirem a distinção entre indivíduos, permitem inferir acerca da origem geográfica;
- A análise AMOVA demonstrou maior variação entre indivíduos do que entre as populações, tal como verificado por outros autores para parte dos marcadores em estudo;
- Por comparação dos genótipos e das frequências alélicas obtidas, verificou-se que as populações em estudo apresentam uma baixa diferenciação e, que se encontram geneticamente próximas, confirmando o papel dos indivíduos guineenses na origem do povo cabo-verdiano, tal como verificado noutros estudos com linhagens maternas.
- A população portuguesa pouco contribuiu para a constituição genética atual da população cabo-verdiana a residir em Portugal, como verificado em diferentes estudos das linhagens paternas na população cabo-verdiana. Este facto sugere a presença de outras populações aquando da colonização inicial do arquipélago, tal como verificado pela análise filogenética;
- Relativamente ao estudo da população guineense a residir em Portugal e comparando com o estudo da população portuguesa, obteve-se um F_{st} próximo de 1, indicativo de uma elevada diferenciação genética entre estas duas populações;
- A análise das distâncias genéticas revelou uma maior proximidade das populações em estudo com a população espanhola do que com a população portuguesa, indicando que a forte presença de espanhóis quer em Cabo Verde quer na Guiné-Bissau, terá contribuído para a constituição genética dessas populações;
- As frequências alélicas obtidas poderão tornar-se úteis noutros estudos populacionais e de evolução, após publicação dos dados nas bases de dados ALFRED (para os marcadores InDel e SNP) e *SNPforID* (para os marcadores SNP apenas).

5. Referências Bibliográficas

Abreu, A., (2011) “Mobilidade e migrações na Guiné-Bissau: dinâmicas históricas e determinantes estruturais”, *Revista Internacional Em Língua Portuguesa - Migrações*, (23), pp.145–161.

Disponível em www.ces.uc.pt/myces/UserFiles/livros/1097_livro_migracoes.pdf

Ahn, S.-M., *et al.* (2009) “The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group”, *Genome Research*, 19(9), pp.1622–9.

Akhteruzzaman, S., *et al.* (2013) “Genetic polymorphism of 30 InDel markers for forensic use in Bangladeshi population”, *Forensic Science International: Genetics Supplement Series*, 4(1), pp.e348–e349, <http://dx.doi.org/10.1016/j.fsigss.2013.10.177>

Amigo, J., *et al.* (2008) “The SNPforID browser: an online tool for query and display of frequency data from the SNPforID project”, *International Journal of Legal Medicine*, 122(5), pp.435–40.

Amorim, A. e Pereira, L. (2005) “Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs”, *Forensic Science International*, 150, pp.17–21.

Andrade, E., (1996) “Cabo Verde: do seu achamento à independência nacional” [Consultado em 18.01.2014]

Disponível em <http://www.governo.cv/>

Applied Biosystems (2012) “SNaPshot ® Multiplex System for SNP genotyping. One system, many applications”. [Consultado em 12.01.2014]

Disponível em <http://www.lifetechnologies.com>

Ardlie, K.G., *et al.* (2002) “Patterns of linkage disequilibrium in the human genome”, *Nature Reviews: Genetics*, 3(4), pp.299–309.

Aronson, J.D., (2005) “DNA fingerprinting on trial: the dramatic early history of a new forensic technique”, *Endeavour*, 29(3), pp.126–31.

Barbaro, A., *et al.* (2012) “Genetic variability of the SNPforID 52-plex identification SNP panel in Italian population samples”, *Forensic Science International: Genetics*, 6, pp.e185–6.

Barcellos, C.J.S. (1899) “Subsídios para a História de Cabo Verde e Guiné”, Academia Real das Ciências de Lisboa, Lisboa. [Consultado em 18.09.2013]

Disponível em <http://books.google.com/?hl=pt-PT>.

- Bastos-Rodrigues, L., *et al.* (2006) “The genetic structure of human populations studied through short insertion-deletion polymorphisms”, *Annals of Human Genetics*, 70(Pt5), pp.658–65.
- Bentley, D.R., *et al.* (2008) “Accurate whole human genome sequencing using reversible terminator chemistry”, *Nature*, 456(7218), pp.53–9.
- Bhangale, T.R., *et al.* (2005) “Comprehensive identification and characterization of diallelic insertion-deletion polymorphisms in 330 human candidate genes”, *Human Molecular Genetics*, 14(1), pp.59–69.
- Bland, J.M. e Altman, D.G. (1995) “Multiple significance tests: the Bonferroni method”, *BMJ*, 310, pp.170.
- Børsting, C., *et al.* (2009) “Validation of a single nucleotide polymorphism (SNP) typing assay with 49 SNPs for forensic genetic testing in a laboratory accredited according to the ISO 17025 standard”, *Forensic Science International: Genetics*, 4(1), pp.34–42.
- Brehm, A., *et al.* (2002) “Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade”, *Annals of Human Genetics*, 66(Pt1), pp.49–60.
- Brookes, A.J. (1999) “The essence of SNPs”, *Gene*, 234(2), pp.177–86.
- Brumfield, R.T., *et al.* (2003) “The utility of single nucleotide polymorphisms in inferences of population history”, *Trends in Ecology & Evolution*, 18(5), pp. 249–256.
- Buckleton, J., Triggs, C.M. e Walsh, S.J. (2005) “Population Genetic Models”, in Buckleton J., Triggs, C.M. e Walsh, S.J. (Eds.) *Forensic DNA Evidence Interpretation*, pp. 65–122, CRC Press, Florida, USA.
- Bulbul, O.P., *et al.* (2009) “Internal validation of 29 autosomal SNP multiplex using a ABI 310 genetic analyser”, *Forensic Science International: Genetics Supplement Series*, 2(1), pp.129–130.
- Butler, J.M. (2005a) “DNA Biology Review” in Butler, J.M. *Forensic DNA Typing: Biology, Technology, and Genetics of the STR Markers*, 2nd ed., pp. 17–31, Elsevier, USA.
- Butler, J.M. (2005b) “STR Population Database Analyses” in Butler, J.M. *Forensic DNA Typing: Biology, Technology and Genetics of STR Markers*, 2nd ed., pp. 473–496, Elsevier, USA.

Butler, J.M. (2005c) “Overview and History of DNA Typing” in Butler, J.M. *Forensic DNA Typing, Biology: Technology and Genetics of STR Markers*, 2nd ed., pp. 1–13, Elsevier, USA.

Butler, J.M. (2010a) “Additional Loci and Nonhuman DNA Testing” in Butler, J.M. *Fundamentals of Forensic DNA Typing*, 2nd ed., pp. 341–362, Academic Press, USA.

Butler, J.M. (2010b) “Basics of DNA Biology and Genetics” in Butler, J.M. *Fundamentals of forensic DNA typing*, 2nd ed., pp. 19–41, Academic Press, USA.

Cá e Sá, I. de F. (2012) “As Comunidades Políticas Transnacionais: O Caso das Associações da Guiné-Bissau em Portugal”, Dissertação de Mestrado em Desenvolvimento e Cooperação Internacional, Instituto Superior de Economia e Gestão, Universidade Técnica de Lisboa, Lisboa.

Cardon, L.R. e Bell, J.I. (2001) “Association study designs for complex diseases”, *Nature Reviews: Genetics*, 2(2), pp.91–9.

Carracedo, A. e Sanchez-Diz, P. (2005) “Forensic DNA Typing Protocols: Forensic DNA-Typing Technologies”, in Carracedo A. (Ed.) *Methods in molecular biology*, Vol. 297, pp. 1–11, Humana Press, Totowa, N.J.

Carreira, A. (1981) “O Tráfico de Escravos nos Rios de Guiné e Ilhas de Cabo Verde (1810-1850) - Subsídios para o seu estudo”, Junta de Investigações Científicas do Ultramar - Centro de Estudos de Antropologia Cultural, Lisboa.

Carreira, A. (1983a) “Documentos para a História das Ilhas de Cabo Verde e ‘Rios da Guiné’ (Séculos XVII e XVIII)”, Gráfica Europam, LDA., Lisboa.

Carreira, A. (1983b) “Migrações nas Ilhas de Cabo Verde”, 2a Edição, Instituto Caboverdeano do Livro, Lisboa.

Carvalho, A. e Pinheiro, M.F. (2013) “Population data of 30 insertion/deletion polymorphisms from a sample taken in the North of Portugal”, *International Journal of Legal Medicine*, 127(1) pp.65–7.

Carvalho, F.A., (2010) “Migração em Cabo Verde, Perfil Nacional 2009”, Organização Internacional para as Migrações, Suíça.

Carvalho, M., *et al.* (2011) “Paternal and maternal lineages in Guinea-Bissau population”, *Forensic Science International: Genetics*, 5(2), pp.114–6.

Carvalho, M. R. (2010) “Single Nucleotide Polymorphisms (SNPs) com Aplicação Forense”, in M. F. Pinheiro (Ed.), *Genética Forense: Perspectivas da Identificação Genética*, pp. 187-197, Edições Universidade Fernando Pessoa, Porto.

Comunidade dos Países de Língua Portuguesa, (2012) “Estatísticas da CPLP. [Consultado em 19.01.2014]

Disponível em http://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_main

Costa, G., *et al.* (2008) “Autosomal SNPs in paternity investigation”, *Forensic Science International: Genetics Supplement Series*, 1(1), pp.507–509.

Costa Francez, P.A. da, *et al.* (2012) “Insertion-deletion polymorphisms--utilization on forensic analysis”, *International Journal of Legal Medicine*, 126(4), pp.491–6.

Dario, P., *et al.* (2009) “SNPs in paternity investigation: The simple future”, *Forensic Science International: Genetics Supplement Series*, 2(1), pp.127–128.

Dario, P. A. P. S. (2010) “SNPs Autossômicos em Identificação Forense”, Dissertação de Mestrado em Medicina Legal e Ciências Forenses, Faculdade de Medicina da Universidade de Lisboa, Lisboa.

Dawson, E., *et al.* (2001) “A SNP Resource for Human Chromosome 22: Extracting Dense Clusters of SNPs From the Genomic Sequence”, *Genome Research*, 11(1), pp.170–178.

dbSNP, Short Genetic Variations [Consultada em 10.12.2013]

Disponível em <http://www.ncbi.nlm.nih.gov/SNP/>

Dereeper, A., *et al.* (2008) “Phylogeny.fr: robust phylogenetic analysis for the non-specialist” *Nucleic Acids Research*, 36, pp.W465–9.

Dereeper, A., *et al.* (2010) “BLAST-EXPLORER helps you building datasets for phylogenetic analysis”, *BMC Evolutionary Biology*, 10(8), pp.8.

Dixon, L., *et al.* (2005) “Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes”, *Forensic Science International*, 154(1), pp.62–77.

Elias, I. e Lagergren, J. (2007) “Fast computation of distance estimators”, *BMC Bioinformatics*, 8, pp.89.

Excoffier, L., *et al.* (1992) “Analysis of Molecular Variance Inferred from Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data”, *Genetics*, 131(2), pp.479–491.

Excoffier, L., *et al.* (2005) “Arlequin ver. 3.0: An Integrated Software Package for Population Genetics Data Analysis”, *Evolutionary Bioinformatics*, 1, pp.47–50.

Fang, R., *et al.* (2009) “Multiplexed SNP detection panels for human identification”, *Forensic Science International: Genetics Supplement Series*, 2(1), pp.538–539.

Felsenstein, J. (1989) “PHYLIP- Phylogeny Inference Package”. [Consultado em 10.02.2014]
Disponível em <http://evolution.genetics.washington.edu/phylip.html>

Fernandes, T., *et al.* (2003) “Genetic differentiation of the Cabo Verde archipelago population analysed by STR polymorphisms”, *Annals of Human Genetics*, 67(Pt 4), pp.340–7.

Ferronha, A.L. (1994) “O Comércio Português de Escravos”, Grupo de Trabalho do Ministério da Educação para as Comemorações dos Descobrimentos Portugueses, Lisboa.

Fondevila, M., *et al.* (2011a) “Forensic performance of insertion–deletion marker systems” *Forensic Science International: Genetics Supplement Series*, 3(1), pp.e443–e444.

Fondevila, M., *et al.* (2011b) “An Assessment of Linkage Between Forensic Markers: Core STRs, Mini-STRs and InDels”, in *Proceedings of the 22nd International Symposium on Human Identification*. [Consultado em 20.09.2013]
Disponível em <http://www.cstl.nist.gov/>

Fondevila, M., *et al.* (2012) “Forensic performance of two insertion-deletion marker assays”, *International Journal of Legal Medicine*, 126(5), pp.725–37.

Fondevila, M., *et al.* (2013) “Revision of the SNPforID 34-plex forensic ancestry test: Assay enhancements, standard reference sample genotypes and extended population studies”, *Forensic Science International: Genetics*, 7(1), pp.63–74.

Freire-Aradas, A., *et al.* (2012) “A new SNP assay for identification of highly degraded human DNA”, *Forensic Science International: Genetics*, 6(3), pp.341–9.

Friis, S.L., *et al.* (2012) “Typing of 30 insertion/deletions in Danes using the first commercial indel kit--Mentype® DIPplex”, *Forensic Science International: Genetics*, 6(2), pp.e72–4.

Frudakis, T., *et al.* (2003) “A classifier for the SNP-based inference of ancestry”, *Journal of Forensic Sciences*, 48(4), pp.771–82.

Fundação Francisco Manuel dos Santos, “Pordata - Base de Dados Portugal Contemporâneo”. [Consultada a 04.01.2014]

Disponível em <http://www.pordata.pt/>

Gascuel, O. (1997) “BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data”, *Molecular Biology and Evolution*, 14(7), pp.685–695.

Gill, P.J., *et al.* (1985) “Forensic application of DNA ‘fingerprints’”, *Nature*, 318(6046), pp.577–579.

Gonçalves, R., *et al.* (2003) “Y-chromosome lineages in Cabo Verde Islands witness the diverse geographic origin of its first male settlers”, *Human Genetics*, 113(6), pp.467–72.

Goodwin, W., *et al.* (2007a) “DNA structure and the genome”, in Goodwin, W., Linacre, A. e Hadi, S., *An Introduction to Forensic Genetics*, pp. 7–15, John Wiley and Sons, Ltd, Inglaterra.

Goodwin, W., *et al.* (2007b) “Single Nucleotide Polymorphisms”, in Goodwin, W., Linacre, A. e Hadi, S., (Eds.) *An Introduction to Forensic Genetics*, pp. 115–123, Wiley & Sons, Inc., Inglaterra.

Goodwin, W., *et al.* (2007c) “Statistical interpretation of STR profiles”, in Goodwin, W., Linacre, A. e Hadi, S. (Eds.) *An Introduction to Forensic Genetics*, pp. 75–84, Wiley & Sons, Inc., Inglaterra.

Guo, S. e Thompson, E. (1992) “Performing the exact test of Hardy-Weinberg proportion for multiple alleles” *Biometrics*, 48(2), pp.361–372.

Hildebrand, C.E., *et al.* (1992) “Mapping the Genome: Informativeness of Polymorphic DNA Markers”, *Los Alamos Science*, (20), pp.100–102.

Hinds, D., *et al.* (2005) “Whole-genome patterns of common DNA variation in three human populations”, *Science*, 307(5712), pp.1072–9.

INE de Cabo Verde, “Portal do Instituto Nacional de Estatística”. [Consultado em 24.01.2014]

Disponível em <http://www.ine.cv/index.aspx>

Instituto Nacional de Estatística da Guiné-Bissau [Consultado em 24.01.2014]

Disponível em <http://www.stat-guinebissau.com/index.htm>

- Jeffreys, A.J., *et al.* (1985a) “Positive identification of an immigration test-case using human DNA fingerprints”, *Nature*, 317(6040), pp.818–819.
- Jeffreys, A.J., *et al.* (1985b) “Hypervariable ‘minisatellite’ regions in human DNA”, *Nature*, 314(6006), pp.67–73.
- Kass, D.H. e Batzer, M.A. (2001) “Genome organization: human”, *Encyclopedia of Life Sciences*
Disponível em <http://onlinelibrary.wiley.com/doi/10.1038/npg.els.0003855/full>
- Kauppi, L., *et al.* (2003) “Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region”, *Human Molecular Genetics*, 12(1), pp.33–40.
- Kidd, J.M., *et al.* (2008) “Mapping and sequencing of structural variation from eight human genomes”, *Nature*, 453(7191), pp.56–64.
- Kidd, K.K., *et al.* (2006) “Developing a SNP panel for forensic identification of individuals”, *Forensic Science International*, 164(1), pp.20–32.
- Kidd, K.K., *et al.* (2012) “Expanding data and resources for forensic use of SNPs in individual identification”, *Forensic Science International: Genetics*, 6(5), pp.646–652.
- Kim, E., *et al.* (2013) “Population data for 30 insertion–deletion markers in a Korean population”, *International Journal of Legal Medicine*, 5, pp.10–11.
- Kim, J.-I., *et al.* (2009) “A highly annotated whole-genome sequence of a Korean individual”, *Nature*, 460(7258), pp.1011–5.
- Kis, Z., *et al.* (2012) “Genome deletion and insertion polymorphisms (DIPs) in the Hungarian population”, *Forensic Science International: Genetics*, 6(5), pp.e125–6.
- Kosoy, R., *et al.* (2009) “Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America”, *Human Mutation*, 30(1), pp.69–78.
- Kuppuswamy, M.N.H., *et al.* (1991) “Single nucleotide primer extension to detect genetic diseases: experimental application to hemophilia B (factor IX) and cystic fibrosis genes”, *Proceedings of the National Academy of Sciences of the United States of America*, 88(4), pp.1143–7.

- Lao, O., *et al.* (2006) “Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry”, *American Journal of Human Genetics*, 78(4), pp.680–90.
- LaRue, B.L., *et al.* (2012) “A validation study of the Qiagen Investigator DIPplex® kit; an INDEL-based assay for human identification”, *International Journal of Legal Medicine*, 126(4), pp.533–40.
- Levy, S., *et al.* (2007) “The diploid genome sequence of an individual human”, *PLoS Biology*, 5(10), pp.e254.
- Ley, T.J., *et al.* (2008) “DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome”, *Nature*, 456(7218), pp.66–72.
- Li, C., *et al.* (2011) “Genetic polymorphism of 29 highly informative InDel markers for forensic use in the Chinese Han population”, *Forensic Science International: Genetics*, 5(1) pp.e27–30.
- Makałowski, W. (2001) “The human genome structure and organization”, *Acta Biochimica Polonica*, 48(3), pp.587–98.
- Martín, P., *et al.* (2013), “Population genetic data of 30 autosomal indels in Central Spain and the Basque Country populations”, *Forensic Science International: Genetics*, 7(2) pp.e27–30.
- Martins, T. e Carvalho, R. (2008) “Conceitos básicos de genética com aplicação à investigação criminal”, in Pinheiro, M. (Eds.) *CSI Criminal*, pp. 83–94, Edições Universidade Fernando Pessoa, Porto.
- Matos, S. (2013) Estudo de novos marcadores genéticos do tipo STR e InDel, Dissertação de Mestrado em Biologia Molecular em Saúde, Escola Superior de Saúde Egas Moniz, Monte da Caparica, Almada.
- Mills, R.E., *et al.* (2006) “An initial map of insertion and deletion (INDEL) variation in the human genome”, *Genome Research*, 16, pp. 1182-1190.
- Mills, R.E., *et al.* (2011) “Natural genetic variation caused by small insertions and deletions in the human genome”, *Genome Research*, 21(6), pp.830–9.
- Montelius, K., *et al.* (2009) “Swedish population data on the SNPforID consortium autosomal SNP-multiplex”, *Forensic Science International: Genetics Supplement Series*, 2(1), pp.344–346.

- Morais, P. (2013) Estudo do DNA mitocondrial de um grupo de imigrantes oriundos de Cabo Verde residentes em Lisboa, Dissertação de Mestrado em Biologia Molecular em Saúde Escola Superior de Saúde Egas Moniz, Monte da Caparica, Almada.
- Mullaney, J. e Mills, R. (2010) “Small insertions and deletions (INDELs) in human genomes” *Human Molecular Genetics*, 19(2), pp.131–136.
- Mullikin, J.C., *et al.* (2000) “An SNP map of human chromosome 22”, *Nature*, 407(6803), pp. 516-20.
- Neuvonen, A.M., *et al.* (2012) “Discrimination power of Investigator DIPplex loci in Finnish and Somali populations”, *Forensic Science International: Genetics*, 6(4), pp.e99–102.
- Nievergelt, C.M., *et al.* (2013) “Inference of human continental origin and admixture proportions using a highly discriminative ancestry informative 41-SNP panel” *Investigative Genetics*, 4(1), pp.13.
- Nussbaum, R., McInnes, R. e Willard, H. (2007a) “Genetic Variation in Individuals and Populations: Mutation and Polymorphisms”, in Nussbaum, R., McInnes, R. e Willard, H. (Eds.) *Thompson & Thompson Genetics in Medicine*, pp. 175–204, Saunders: Elsevier, Canada.
- Nussbaum, R. McInnes, R. e Willard, H. (2007b) “The Human Genoma: Gene Structure and Function”, in Nussbaum, R., McInnes, R. e Willard, H. (Eds.) *Thompson & Thompson Genetics in Medicine*, 7 Edition., pp. 25–30, Saunders: Elsevier, Canada.
- Page, R.D.M. (1996) “TREEVIEW: An application to display phylogenetic trees on personal computers”, *Computer Applications in the Biosciences*, 12, pp.357–358.
- PAIGC (1974) “História da Guiné e Ilhas de Cabo Verde”, Afrontamento, Porto.
- Pasternak, J.J. (2005) “Human Population Genetics”, in Pasternak, J.J. (Eds.) *An Introduction to Human Molecular Genetics: Mechanisms of Inherited Diseases*, 2 Ed, pp. 229–286, Wiley & Sons, Inc., Canada.
- Pena, H.B. e Pena, S.D J. (2012) “Automated Genotyping of a Highly Informative Panel of 40 Short Insertion-Deletion Polymorphisms Resolved in Polyacrylamide Gels for Forensic Identification and Kinship Analysis”, *Transfusion Medicine and Hemotherapy*, 39(3), pp.211–216.
- Pepinski, W., *et al.* (2013) “Population genetics of 30 INDELs in populations of Poland and Taiwan”, *Molecular Biology Reports*, 40(7), pp.4333–8.

Pereira, R., *et al.* (2008) “Genetic characterization of 52 autosomal SNPs in the Portuguese population”, *Forensic Science International: Genetics Supplement Series*, 1, pp.358–360.

Pereira, R. e Gusmão, L. (2012) “Capillary electrophoresis of 38 noncoding biallelic mini-Indels for degraded samples and as complementary tool in paternity testing”, in Alonso, A. *DNA Electrophoresis Protocols for Forensic Genetics, Methods in Molecular Biology*, 830, pp.141–157, Humana Press, Totowa, N.J.

Pereira, R., *et al.* (2009a) “A new multiplex for human identification using insertion/deletion polymorphisms”, *Electrophoresis*, 30(21), pp.3682–90.

Pereira, R., *et al.* (2009b) “Insertion/deletion polymorphisms: A multiplex assay and forensic applications”, *Forensic Science International: Genetics Supplement Series*, 2(1), pp.513–515.

Pereira, R., *et al.* (2012) “Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing”, *PloS One*, 7(1), pp.e29684.

Phillips, C. (2005) “Forensic DNA Typing Protocols: Using Database for Developing SNP Markers of Forensic Interest”, in Carracedo A. (Eds.) *Methods in molecular biology*, pp. 83–106, Humana Press, Totowa, N.J.

Phillips, C., *et al.* (2007) “Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs”, *Forensic Science International: Genetics*, 1(3-4), pp. 273-80.

Pimenta, J. e Pena, S. (2010) “Efficient human paternity testing with a panel of 40 short insertion-deletion polymorphisms”, *Genetics and Molecular Research*, 9(1), pp.601–607.

Pinheiro, M.F., (2008) “A perícia em genética e biologia forense - criminalística biológica”, in Pinheiro, M.F. *CSI Criminal*, pp. 11–40, Edições Universidade Fernando Pessoa, Porto.

Pinheiro, M.F. (2010) “Algumas perspectivas da identificação genética”, in Pinheiro, M.F. *Genética forense: perspectivas da identificação genética*, pp. 17–78, Edições Universidade Fernando Pessoa, Porto.

Poulsen, L., *et al.* (2011) “Typing of Amerindian Kichwas and Mestizos from Ecuador with the SNPforID multiplex” *Forensic Science International: Genetics*, 5(4), pp.e105–7.

QIAGEN (2010), “Population data for Investigator DIPplex”.

Disponível em <http://b2b.qiagen.com/products/catalog/assay-technologies/investigator-diplex-kit#resources>

- QIAGEN, (2013) “Investigator ® DIPplex Handbook Sample & Assay Technologies”. Disponível em <http://b2b.qiagen.com/products/catalog/assay-technologies/investigator-dipplex-kit#resources>.
- Romanini, C., *et al.* (2012) “Typing short amplicon binary polymorphisms: supplementary SNP and Indel genetic information in the analysis of highly degraded skeletal remains”, *Forensic Science International: Genetics*, 6(4), pp.469–76.
- Romualdi, C., *et al.* (2002) “Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms”, *Genome Research*, 12(4), pp.602–12.
- Rosa, A., *et al.* (2004) “MtDNA profile of West Africa Guineans: towards a better understanding of the Senegambia region”, *Annals of Human Genetics*, 68(Pt 4), pp.340–52.
- Rosa, A., *et al.* (2007) “Y-chromosomal diversity in the population of Guinea-Bissau: a multiethnic perspective”, *BMC Evolutionary Biology*, 7, pp. 124.
- Rosenberg, N.A., *et al.* (2005) “Clines, clusters, and the effect of study design on the inference of human population structure”, *PLoS Genetics*, 1(6), pp.e70.
- Rosenberg, N.A., *et al.* (2002) “Genetic structure of human populations”, *Science* (New York, N.Y.) 298(5602), pp.2381–5.
- Sanchez, J.J., *et al.* (2006) “A multiplex assay with 52 single nucleotide polymorphisms for human identification”, *Electrophoresis*, 27(9), pp.1713–24.
- Sanchez, J.J., *et al.* (2008) “Forensic typing of autosomal SNPs with a 29 SNP-multiplex--results of a collaborative EDNAP exercise”, *Forensic Science International: Genetics*, 2(3), pp.176–83.
- Santos, C., *et al.* (2011) “A study of East Timor variability using the SNPforID 52-plex SNP panel”, *Forensic Science International: Genetics*, 5(1), pp.e25–6.
- Santos, N.P.C., *et al.* (2010) “Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel”, *Human Mutation*, 31(2), pp.184–90.
- Schuster, S.C., *et al.* (2014) “Population genetics of insertion-deletion polymorphisms in South Koreans using Investigator DIPplex kit”, *Forensic Science International: Genetics*, 8(1), pp.80–3.

Serviço de Estrangeiros e Fronteiras (2013) “Relatório de imigração fronteiras e asilo 2012” pp. 16–17, Oeiras.

Disponível em: <http://www.sef.pt/portal/v10/PT/asp/page.aspx>

Sharafi Farzad M., *et al.* (2013) “Analysis of 49 autosomal SNPs in three ethnic groups from Iran: Persians, Lurs and Kurds”, *Forensic Science International: Genetics*, 7(4), pp.471–3.

Silva, C.V. da, *et al.* (2013) “Genetic portrait of south Portugal population with InDel markers”, *Forensic Science International: Genetics*, 7(4), pp.e101–3.

Slatkin, M. e Excoffier, L. (1996) “Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm”, *Heredity*, 76 (Pt4), pp.377–83.

Sobrino, B., Brión, M. e Carracedo, A. (2005) “SNPs in forensic genetics: a review on SNP typing methodologies”, *Forensic Science International*, 154(2-3), pp.181–94.

Sokolov, B.P. (1990) “Primer extension technique for the detection of single nucleotide in genomic DNA”, *Nucleic Acids Research*, 18(12), pp.3671.

Souto, L. (2010) “Interesse dos microssatélites na identificação genética”, in Pinheiro, M.F. T (Eds.) *Genética Forense: Perspectivas da identificação genética*, pp. 79 – 102, Edições Universidade Fernando Pessoa, Porto.

Spinola, H., *et al.* (2002) “Distribution of HLA alleles in Portugal and Cabo Verde. Relationships with the slave trade route”, *Annals of Human Genetics*, 66(Pt4), pp.285–96.

SNPforID browser. [Consultado em 13.12.2013]

Disponível em <http://spsmart.cesga.es/snpforid.php>

Syvänen, A.C., *et al.* (1990) “A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E”, *Genomics*, 8(4) pp.684–92.

Takezaki, N. *et al.* (2010) “POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface” *Molecular Biology and Evolution*, 27 pp.747-752.

Disponível em

<http://www.med.kagawa-u.ac.jp/~genomelb/takezaki/poptree2/index.html>

Tereba, A. (1999) “Tools for Analysis of Population Statistics”, *Profiles in DNA*, pp.14–16.

Tomas, C., *et al.* (2013) “Analysis of 49 autosomal SNPs in an Iraqi population”, *Forensic Science International: Genetics*, 7(1), pp.198–9.

- Turrina, S., *et al.* (2011) “Forensic evaluation of the Investigator DIPplex typing system”, *Forensic Science International: Genetics Supplement Series*, 3(1) pp.e331–e332.
- Vallone, P.M., *et al.* (2006) “The evaluation of an autosomal SNP 12-plex assay”, *International Congress Series*, 1288, pp.61–63.
- Vallone, P.M., *et al.* (2005) “Allele frequencies for 70 autosomal SNP loci with U.S. Caucasian, African-American, and Hispanic samples”, *Forensic Science International*, 149(2-3), pp.279–86.
- Vullo, C., *et al.* (2011) “Population data for 38 autosomal insertion/deletion (InDels) and 50 SNPS polymorphisms in Argentinean population”, *Forensic Science International: Genetics Supplement Series*, 3(1), pp.e419–e420.
- Walsh, P.S., *et al.* (1991) “Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material”, *BioTechniques*, 10(4), pp.506–13.
- Wang, J., *et al.* (2008) “The diploid genome sequence of an Asian individual”, *Nature*, 456(7218), pp.60–5.
- Weber, J.L., *et al.* (2002) “Human diallelic insertion/deletion polymorphisms”, *American Journal of Human Genetics*, 71(4), pp.854–62.
- Wei, Y.-L., *et al.* (2014) “A validation study of a multiplex INDEL assay for forensic use in four Chinese populations”, *Forensic Science International: Genetics*, 9, pp.e22–5.
- Westen, A., *et al.* (2012) “Combining results of forensic STR kits: HDplex validation including allelic association and linkage testing with NGM and Identifiler loci”, *International Journal of Legal Medicine*, 126(5), pp.781–9.
- Wheeler, D.A., *et al.* (2008) “The complete genome of an individual by massively parallel DNA sequencing”, *Nature*, 452(7189), pp.872–6.
- Whittle, M.R., *et al.* (2009) “Paternity investigation experience with a 40 autosomal SNP panel”, *Forensic Science International: Genetics Supplement Series*, 2(1), pp.149–150.
- Yang, N., *et al.* (2005) “Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine”, *Human Genetics*, 118(3-4), pp.382–92.

Zaumsegel, D., *et al.* (2013) “A 21 marker insertion deletion polymorphism panel to study biogeographic ancestry”, *Forensic Science International: Genetics*, 7(2), pp.305–12.

Zidkova, A., *et al.* (2013) “Application of the new insertion-deletion polymorphism kit for forensic identification and parentage testing on the Czech population”, *International Journal of Legal Medicine*, 127(1), pp.7–10.

6. Anexos

6.1. Percentagens de variações máximas e mínimas da análise intra- e interpopulacional

Variação máxima e mínima observada e os respetivos *loci*, detetadas entre os três grupos populacionais estudados: Guiné-Bissau (GB) e Cabo Verde (CV), Guiné-Bissau e Portugal (PT), Cabo Verde e Portugal.

% Variação		Entre Populações	Entre Indivíduos dentro Populações	Entre Indivíduos	Fst
GB vs CV	Máxima	HLD124 8,44567	HLD97 24,47047	A37 115,06166	HLD124 0,08446
	Mínima	A51 -0,46675	A36 -14,88765	HLD97 75,70845	A10 -0,00461
GB vs PT	Máxima	A13 28,31578	HLD97 30,26770	A37 114,34924	A13 0,28316
	Mínima	HLD93 -0,54823	A29 -17,10073	HLD101 68,58360	HLD93 -0,00548
CV vs PT	Máxima	HLD70 20,69187	HLD97 16,29021	A12 109,03536	HLD58 0,20227
	Mínima	A06 -0,30141	HLD111 -9,11153	HLD70 79,90286	A06 -0,00301

6.2. Valores de Fst obtidos por locus

Valores de diferenciação genética (Fst) obtidos aquando da comparação da população guineense (GB) com a população cabo-verdiana (CV), da população guineense com a população portuguesa (PT) e da população cabo-verdiana com a população portuguesa, para os 50 *loci* em estudo e as respetivas médias. Os valores de *p* significativos ($p < 0,05$) encontram-se realçados.

<i>Locus</i>	GB vs CV		GB vs PT		CV vs PT	
	Fst	<i>p</i>	Fst	<i>p</i>	Fst	<i>p</i>
HLD77	0,00028	0,46900	0,04715	0,00337	0,02654	0,00225
HLD45	-0,00327	1,00000	0,03393	0,01006	0,02715	0,00275
HLD131	0,00252	0,32177	0,14339	0,00000	0,10799	0,00000
HLD70	0,01150	0,05318	0,25552	0,00000	0,20692	0,00000
HLD6	-0,00358	1,00000	-0,00456	1,00000	-0,00270	1,00000
HLD111	0,00592	0,11424	0,09189	0,00000	0,04679	0,00000
HLD58	0,01666	0,02537	0,27065	0,00000	0,20227	0,00000
HLD56	0,01485	0,03162	0,06413	0,00031	0,01394	0,01769
HLD118	0,01555	0,02993	0,05657	0,00044	0,00996	0,04481
HLD92	-0,00369	1,00000	0,03676	0,00525	0,03965	0,00006
HLD93	0,00436	0,17542	-0,00548	1,00000	0,00321	0,20485
HLD99	-0,00212	1,00000	0,04085	0,00275	0,04112	0,00019
HLD88	0,00005	0,47413	0,05556	0,00119	0,03316	0,00050
HLD101	0,03205	0,00237	0,25904	0,00000	0,13503	0,00000
HLD67	0,00281	0,27440	0,00280	0,33752	-0,00271	1,00000
HLD83	0,01248	0,06268	0,06814	0,00012	0,01810	0,01500
HLD114	0,00953	0,08449	0,20676	0,00000	0,13424	0,00000
HLD48	0,04232	0,00037	0,12818	0,00000	0,02288	0,00319
HLD124	0,08446	0,00000	0,32452	0,00000	0,10257	0,00000
HLD122	-0,00342	1,00000	-0,00396	1,00000	-0,00275	1,00000
HLD125	0,02743	0,00381	0,17696	0,00000	0,07108	0,00000
HLD64	0,00265	0,30009	0,11415	0,00006	0,07781	0,00000
HLD81	0,00760	0,10136	0,07069	0,00025	0,02649	0,00256
HLD136	0,00629	0,12505	0,16361	0,00000	0,11457	0,00000
HLD133	0,00076	0,38920	0,02007	0,03206	0,00531	0,14404
HLD97	-0,00179	0,70116	0,00861	0,17392	0,00159	0,34514
HLD40	0,00171	0,32083	0,11436	0,00000	0,07768	0,00000
HLD128	0,02020	0,01531	0,19016	0,00000	0,09103	0,00000
HLD39	0,00334	0,24959	0,03637	0,00662	0,01198	0,03468
HLD84	-0,00395	1,00000	0,00221	0,33815	0,00332	0,20491
A01	0,01399	0,03431	0,00275	0,33633	-0,00112	0,74322
A27	-0,00248	1,00000	0,07046	0,00006	0,05597	0,00000
A51	-0,00467	1,00000	0,00847	0,17179	0,00871	0,07555
A10	-0,00461	1,00000	0,04330	0,00512	0,03752	0,00050

Continuação - Valores de diferenciação genética (Fst) obtidos aquando da comparação da população guineense (GB) com a população cabo-verdiana (CV), da população guineense com a população portuguesa (PT) e da população cabo-verdiana com a população portuguesa, para os 50 *loci* em estudo e as respetivas médias. Os valores de *p* significativos ($p < 0,05$) encontram-se realçados.

<i>Locus</i>	GB vs CV		GB vs PT		CV vs PT	
	Fst	<i>p</i>	Fst	<i>p</i>	Fst	<i>p</i>
A49	0,00095	0,41970	0,00853	0,17192	-0,00129	0,70697
A08	0,00123	0,40026	0,01157	0,11574	0,00039	0,40420
A06	0,01169	0,07318	0,00595	0,20916	-0,00301	1,00000
A12	-0,00341	1,00000	-0,00430	1,00000	-0,00253	1,00000
A14	0,00278	0,28415	0,03136	0,01219	0,00927	0,05993
A50	0,00489	0,23666	0,01476	0,09955	0,05426	0,00012
A41	0,00422	0,21585	0,04533	0,00175	0,01553	0,01719
A13	0,02600	0,00337	0,28316	0,00000	0,16479	0,00000
A36	0,00416	0,16798	0,07101	0,00075	0,03715	0,00012
A38	-0,00181	0,78134	0,05520	0,00087	0,08263	0,00000
A37	-0,00222	1,00000	-0,00276	1,00000	0,00299	0,22910
A29	0,00541	0,15461	0,06457	0,00012	0,02696	0,00200
A16	0,00493	0,17773	0,18376	0,00000	0,14216	0,00000
A5	0,02190	0,01269	0,07623	0,00000	0,01377	0,02393
A4	0,00835	0,10636	0,13133	0,00000	0,06581	0,00000
A35	-0,00193	0,92326	0,04514	0,00187	0,03183	0,00106
Média	0,00758		0,08486		0,04690	

6.3. Distâncias Genômicas entre os 50 marcadores em estudo

Distâncias em nucleótidos (nt) entre os diferentes marcadores estudados, separados por cromossoma.

Cromossoma 1							
dbSNP	Localização (nt)	Locus	A01	HLD40	HLD39	HLD128	A51
rs1490413	4 037 521	A01	0	*	*	*	*
rs2307956	54 718 192	HLD40	50 680 671	0	*	*	*
rs17878444	92 237 892	HLD39	88 200 371	37 519 700	0	*	*
rs2307924	194 878 347	HLD128	190 840 826	140 160 155	102 640 455	0	*
rs891700	236 923 075	A51	232 885 554	182 204 883	144 685 183	42 044 728	0
Cromossoma 2							
			HLD48	HLD45	A38		
rs28369942	100 081 561	HLD48	0	*	*		
rs2307959	169 800 273	HLD45	69 718 712	0	*		
rs907100	239 850 329	A38	139 768 768	70 050 056	0		
Cromossoma 3							
			HLD133	A37			
rs2067235	41 136 221	HLD133	0	*			
rs1355366	192 127 021	A37	150 990 800	0			
Cromossoma 4							
			A04	HLD56			
rs2046361	10 719 942	A04	0	*			
rs2308292	107 889 773	HLD56	97 169 831	0			
Cromossoma 5							
			A05	HLD64	HLD58	HLD67	
rs717302	2 932 133	A05	0	*	*	*	
rs1610935	66 214 500	HLD64	63 282 367	0	*	*	
rs1610937	76 745 067	HLD58	73 812 934	10 530 567	0	*	
rs1305056	155 662 256	HLD67	152 730 123	89 447 756	78 917 189	0	

Continuação - Distâncias em nucleótidos (nt) entre os diferentes marcadores estudados, separados por cromossoma.

Cromossoma 6					
		A06	HLD70		
rs1029047	1 080 939	A06	0	*	
rs2307652	97 458 121	HLD70	96 377 182	0	
Cromossoma 7					
			HLD81	HLD77	HLD131 A41
rs17879936	95 047 150	HLD81	0	*	*
rs1611048	110 939 987	HLD77	15 892 837	0	*
rs1611001	154 404 562	HLD131	59 357 412	43 464 575	*
rs737681	154 850 085	A41	59 802 935	43 910 098	445 523 0
Cromossoma 8					
		A08	HLD83		HLD84
rs763869	1 363 017	A08	0	*	*
rs2308072	19 089 779	HLD83	17 726 762	0	*
rs3081400	119 947 801	HLD84	118 584 784	100 858 022	0
Cromossoma 9					
			HLD88	A35	
rs8190570	98 997 911	HLD88	0	*	
rs1463729	122 257 493	A35	23 259 582	0	
Cromossoma 11					
			HLD92	A36	
rs17174476	102 479 418	HLD92	0	*	
rs2076848	134 205 198	A36	31 725 780	0	
Cromossoma 12					
		A12	HLD93	A27	
rs2107612	741 262	A12	0	*	*
rs2307570	94 675 906	HLD93	93 934 644	0	*
rs2111980	104 830 721	A27	104 089 459	10 154 815	0

Continuação - Distâncias em nucleótidos (nt) entre os diferentes marcadores estudados, separados por cromossoma.

Cromossoma 13				
			A13	HLD97
rs1886510	20 172 700	A13	0	*
rs17238892	31 328 384	HLD97	11 155 684	0
Cromossoma 14				
			A14	HLD99
rs1454361	23 840 960	A14	0	*
rs2308163	58 050 081	HLD99	34 209 121	0
Cromossoma 15				
			A50	HLD101
rs8037429	51 332 965	A50	0	*
rs2307433	89 864 316	HLD101	38 531 351	0
Cromossoma 16				
			A16	HLD6
rs729172	5 606 490	A16	0	*
rs1610905	55 691 830	HLD6	50 085 340	0
Cromossoma 20				
			HLD118	A49
rs16438	25 278 465	HLD118	0	*
rs1005533	40 172 539	A49	14 894 074	0

6.4. Resultados da Análise de LD Entre os 50 Marcadores em Estudo

Tabela representativa dos valores de significância obtidos: a cinzento claro “-” não significativo ($p>0,05$), a cinzento “+” não significativo após correção de Bonferroni ($0,00004<p<0,05$) e a cinzento escuro “+” significativo ($p<0,00004$).

